SYSTEM MCQs COLLECTION

NSCT Prep

Free MCQ Practice for NSCT Test Preparation

AI / Machine Learning & Data Analytics

1530 Multiple Choice Questions

nsctprep.dev

This dataset is created and compiled by Muhammad Abdullah Awais

© 2026 NSCT Prep. All rights reserved.

Easy Questions

510 questions

Q1. What does AI stand for?

A. Advanced Iteration
B. Artificial Intelligence
C. Automated Integration
D. Applied Informatics

Answer: B

Q2. Which of the following is a type of machine learning?

A. Manual Learning
B. Supervised Learning
C. Static Learning
D. Hardware Learning

Answer: B

Q3. Who is considered the father of Artificial Intelligence?

A. John McCarthy
B. Charles Babbage
C. Alan Turing
D. Tim Berners-Lee

Answer: A

Q4. What is the primary goal of machine learning?

A. To design efficient computer hardware
B. To replace human decision-making entirely
C. To write self-modifying operating systems
D. To enable computers to learn from data

Answer: D

Q5. Which of these is an example of AI in daily life?

A. Saving a file to disk
B. Using a basic calculator
C. Printing a text document
D. Spam email filtering

Answer: D

Q6. Data Analytics primarily deals with:

A. Building and assembling computer hardware
B. Writing low-level operating system code
C. Examining data sets to draw conclusions
D. Designing graphical user interface layouts

Answer: C

Q7. What does ML stand for in the context of AI?

A. Machine Learning
B. Meta Language
C. Memory Linking
D. Macro Logic

Answer: A

Q8. Which is NOT a type of machine learning?

A. Reinforcement Learning
B. Compiled Learning
C. Supervised Learning
D. Unsupervised Learning

Answer: B

Q9. A chatbot is an example of:

A. Artificial Intelligence
B. Database Management
C. Operating System Design
D. Computer Architecture

Answer: A

Q10. What is a dataset in data analytics?

A. A sequential optimization algorithm
B. A hardware peripheral device
C. A compiled programming language
D. A collection of related data points

Answer: D

Q11. What is a scalar in linear algebra?

A. A higher-order tensor
B. A multi-row data matrix
C. A single numerical value
D. A multi-element vector

Answer: C

Q12. What is the mean of the numbers 2, 4, 6, 8, 10?

A. 7
B. 8
C. 5
D. 6

Answer: D

Q13. A matrix with equal rows and columns is called:

A. Rectangular matrix
B. Diagonal matrix
C. Square matrix
D. Identity matrix

Answer: C

Q14. What does probability measure?

A. The computational speed of an algorithm
B. The overall size of a given dataset
C. The likelihood of an event occurring
D. The hardware accuracy of a processor

Answer: D

Q15. The derivative of a constant is:

A. 1
B. The constant itself
C. 0
D. Infinity

Answer: C

Q16. What is the dot product of vectors [1,2] and [3,4]?

A. 7
B. 14
C. 11
D. 10

Answer: C

Q17. What is the median of {1, 3, 5, 7, 9}?

A. 3
B. 5
C. 4
D. 7

Answer: B

Q18. An identity matrix has:

A. All entries set to the value of zero
B. Entries filled with random real numbers
C. 1s on the diagonal and 0s elsewhere
D. All entries set to the value of one

Answer: C

Q19. The range of a probability value is:

A. -1 to 1
B. Negative infinity to positive infinity
C. 0 to 1
D. 0 to 100

Answer: C

Q20. What is the transpose of a matrix?

A. The matrix is numerically inverted
B. The matrix entries are all squared
C. The matrix is multiplied by itself
D. Rows and columns are interchanged

Answer: D

Q21. Which Python library is most commonly used for numerical computations in AI?

A. Tkinter
B. Flask
C. NumPy
D. Django

Answer: C

Q22. What does pandas primarily provide?

A. Network socket programming APIs
B. Data structures for data analysis
C. Web development framework tools
D. Game development engine support

Answer: B

Q23. Which library is used for plotting graphs in Python?

A. Pandas
B. NumPy
C. Matplotlib
D. Scikit-learn

Answer: C

Q24. How do you import NumPy with an alias?

A. include numpy
B. import np
C. import numpy as num
D. import numpy as np

Answer: D

Q25. What is a Jupyter Notebook?

A. An interactive computing environment for code and visualizations
B. A production web server for hosting static site content
C. A relational database for storing structured table records
D. A lightweight plain text editor without any execution support

Answer: A

Q26. Which function creates a NumPy array?

A. np.make()
B. np.list()
C. np.array()
D. np.create()

Answer: C

Q27. What does df.head() do in pandas?

A. Sorts the DataFrame by column values
B. Returns the last 5 rows of the frame
C. Deletes the first row of the DataFrame
D. Returns the first 5 rows of the DataFrame

Answer: D

Q28. Which library provides machine learning algorithms in Python?

A. Pillow
B. Scikit-learn
C. Seaborn
D. Matplotlib

Answer: B

Q29. What data type does pandas use for tabular data?

A. List
B. Dictionary
C. DataFrame
D. Array

Answer: C

Q30. Which operator is used for element-wise multiplication in NumPy?

A. *
B. //
C. &
D. @

Answer: A

Q31. What is data preprocessing?

A. Deploying trained models into production systems
B. Writing detailed summary reports on raw findings
C. Transforming raw data into a clean and usable format
D. Creating interactive charts to visualize raw data

Answer: C

Q32. What is a missing value in a dataset?

A. A data point that equals exactly zero
B. A data point that is absent or null
C. An extreme statistical outlier point
D. A strongly negative numerical value

Answer: B

Q33. What is data normalization?

A. Duplicating data across tables
B. Sorting data in alphabetical order
C. Scaling data to a standard range
D. Removing all data from storage

Answer: C

Q34. Which of these is a common data format?

A. SYS
B. EXE
C. CSV
D. DLL

Answer: C

Q35. What does data cleaning involve?

A. Making the dataset physically larger in size
B. Adding more synthetic records to the data
C. Encrypting data columns for secure storage
D. Removing errors and inconsistencies from data

Answer: D

Q36. What is a categorical variable?

A. A variable that can only hold binary data
B. A variable with randomly assigned values
C. A variable with continuous numerical values
D. A variable with discrete categories or labels

Answer: D

Q37. What is a data source?

A. A trained machine learning model file
B. An interactive data visualization chart
C. A computational optimization algorithm
D. The origin from which data is collected

Answer: D

Q38. What is the purpose of removing duplicates?

A. To add new input features
B. To change column data types
C. To eliminate redundant records
D. To increase the dataset size

Answer: C

Q39. What is a structured dataset?

A. Continuous raw audio waveforms
B. Data organized in rows and columns
C. Unorganized raw free-form text
D. Unstructured pixel image data

Answer: B

Q40. Which pandas method checks for null values?

A. fillna()
B. isnull()
C. notnull()
D. dropna()

Answer: B

Q41. What does EDA stand for?

A. Extended Data Algorithm
B. External Data Analysis
C. Effective Data Assessment
D. Exploratory Data Analysis

Answer: D

Q42. Which chart is used to show the distribution of a single variable?

A. Line chart
B. Scatter plot
C. Histogram
D. Pie chart

Answer: C

Q43. What does a box plot display?

A. The five-number summary: min, Q1, median, Q3, max
B. Only the arithmetic mean of the distribution values
C. Only the overall range from minimum to maximum
D. Only the mode or most frequent value observed

Answer: A

Q44. Which pandas method gives summary statistics?

A. describe()
B. info()
C. head()
D. shape

Answer: A

Q45. A scatter plot shows the relationship between:

A. Two numerical variables
B. Categories data only
C. Time series data only
D. Textual data only

Answer: D

Q46. What is the purpose of a correlation matrix?

A. To clean and preprocess the raw input data
B. To deploy and manage software applications
C. To train and evaluate machine learning models
D. To show relationships between multiple variables

Answer: D

Q47. What does df.info() display in pandas?

A. Only the numerical summary statistics values
B. Only the count of missing or null values
C. Column names, data types, and non-null counts
D. Only the first five rows of the DataFrame

Answer: C

Q48. A bar chart is best used for:

A. Showing data distributions
B. Comparing quantities across categories
C. Displaying feature correlations
D. Plotting time series trends

Answer: D

Q49. What is the mode of a dataset?

A. The full range of the values
B. The most frequently occurring value
C. The middle value of the data
D. The average value of the data

Answer: C

Q50. Which plot is best for showing proportions of a whole?

A. Histogram
B. Pie chart
C. Scatter plot
D. Box plot

Answer: B

Q51. In supervised learning, the model learns from:

A. Labeled data with known outputs
B. Randomly generated fake data
C. Unlabeled unstructured data only
D. No data of any kind at all

Answer: A

Q52. Linear regression is used for:

A. Classification tasks only
B. Reducing dataset dimensionality
C. Predicting continuous numerical values
D. Clustering data points

Answer: D

Q53. What is classification in supervised learning?

A. Grouping similar data together
B. Predicting a discrete category or class
C. Predicting a continuous output value
D. Reducing feature dimensions

Answer: C

Q54. Which algorithm is commonly used for classification?

A. K-Means method
B. Decision Tree
C. DBSCAN method
D. PCA method

Answer: D

Q55. What is the target variable in supervised learning?

A. The output the model predicts
B. An internal training weight
C. A tunable hyperparameter
D. An individual input feature

Answer: A

Q56. Logistic regression is used for:

A. Dimensionality reduction
B. Linear regression
C. Binary classification
D. Clustering

Answer: C

Q57. What is overfitting?

A. Model performs well on training data but poorly on new data
B. Model is too simple and has very few learned parameters
C. Model performs poorly on both training and testing data
D. Model has absolutely no trainable weight parameters

Answer: A

Q58. K-Nearest Neighbors (KNN) classifies based on:

A. A low-rank matrix decomposition factorization
B. The steepest gradient descent optimization step
C. An entirely random selection of a class label
D. The majority class among the K nearest data points

Answer: D

Q59. What is a feature in machine learning?

A. The model architecture itself
B. The optimization training algorithm
C. An input variable used for prediction
D. The output target variable label

Answer: C

Q60. What is underfitting?

A. Model has an excessive number of parameters
B. Model has been trained for far too many epochs
C. Model is too simple to capture patterns in data
D. Model has memorized all the training data points

Answer: C

Q61. What is ensemble learning?

A. Using a single standalone model for output
B. Selecting the most relevant input features
C. Combining multiple models to improve prediction
D. Cleaning and preprocessing raw input data

Answer: C

Q62. Random Forest is an ensemble of:

A. Neural networks
B. Linear models
C. Decision trees
D. SVMs

Answer: C

Q63. What is bagging in ensemble learning?

A. Training models on random subsets of data with replacement
B. Cleaning and imputing missing data point values
C. Training a single model on the complete full dataset
D. Removing irrelevant features from the training data

Answer: A

Q64. What is boosting?

A. Sequentially training models where each focuses on previous errors
B. Reducing the overall size of the training dataset
C. Engineering new features from the existing raw data
D. Training all individual models simultaneously in parallel

Answer: A

Q65. The final prediction in a classification ensemble is typically made by:

A. Using only the first model
B. Random class selection
C. Majority voting
D. Using the weakest model

Answer: C

Q66. Which of these is a popular boosting algorithm?

A. DBSCAN
B. K-Means
C. PCA
D. XGBoost

Answer: D

Q67. How does Random Forest handle overfitting?

A. By removing all features from the training input
B. By using a single very deep decision tree only
C. By only adding more raw data samples to train
D. By averaging predictions from many diverse trees

Answer: D

Q68. What is a base learner in ensemble methods?

A. A tunable model hyperparameter value
B. An individual model within the ensemble
C. A data preprocessing pipeline step
D. The final aggregated combined model output

Answer: B

Q69. AdaBoost stands for:

A. Automated Boosting
B. Adaptive Boosting
C. Advanced Boosting
D. Additional Boosting

Answer: B

Q70. In Random Forest, what is feature randomness?

A. Features are randomly generated from scratch each time
B. Each tree considers a random subset of features at each split
C. Features are removed permanently from the entire dataset
D. All decision trees use all available features every time

Answer: A

Q71. Unsupervised learning works with:

A. Unlabeled data
B. Numerical data only
C. Textual data only
D. Labeled output data

Answer: A

Q72. K-Means is a type of:

A. Regression algorithm
B. Clustering algorithm
C. Classification algorithm
D. Sorting algorithm

Answer: B

Q73. What does clustering do?

A. Classifies data with known labels
B. Generates entirely new data points
C. Predicts continuous numerical values
D. Groups similar data points together

Answer: D

Q74. PCA stands for:

A. Principal Component Analysis
B. Primary Cluster Algorithm
C. Probability Calculation Approach
D. Pattern Classification Analysis

Answer: A

Q75. In K-Means, what is a centroid?

A. A class target label
B. An extreme outlier point
C. The center point of a cluster
D. A single input feature

Answer: C

Q76. What is dimensionality reduction?

A. Adding more derived features to expand the feature space
B. Removing all data from the storage system entirely
C. Reducing the number of features while preserving important information
D. Increasing the overall size of the raw training dataset

Answer: C

Q77. Which is an example of unsupervised learning?

A. Predicting house prices
B. Labeled image classification
C. Customer segmentation
D. Email spam detection

Answer: C

Q78. Association rule mining is used to:

A. Find relationships between items in transactions
B. Classify documents into predefined categories
C. Predict future numerical stock price values
D. Train multi-layer deep neural network models

Answer: A

Q79. What does the 'K' in K-Means represent?

A. The number of iterations
B. The number of features
C. The learning rate
D. The number of clusters

Answer: D

Q80. What is an unsupervised learning application?

A. Labeled prediction
B. Anomaly detection
C. Supervised classification
D. Regression

Answer: B

Q81. Accuracy is defined as:

A. Number of correct predictions divided by total predictions
B. Only the count of true negative classification results
C. The total number of features used in the model
D. Only the count of true positive classification results

Answer: A

Q82. What is a confusion matrix?

A. A Python data visualization chart library
B. A type of multi-layer deep neural network
C. A preprocessing data cleaning utility tool
D. A table showing predicted vs actual classifications

Answer: D

Q83. What is precision in classification?

A. The recall evaluation metric
B. Overall model accuracy metric
C. True Positives divided by the total count
D. True Positives / (True Positives + False Positives)

Answer: A

Q84. What is recall (sensitivity)?

A. True Positives / (True Positives + False Positives)
B. True Positives / (True Positives + False Negatives)
C. The specificity metric
D. The overall accuracy metric

Answer: D

Q85. Mean Squared Error (MSE) is used to evaluate:

A. Regression models
B. Classification models only
C. Clustering algorithms
D. Association rule mining

Answer: A

Q86. What is a training set used for?

A. Deploying the model to production
B. Training the model to learn patterns
C. Evaluating the final performance
D. Tuning the model hyperparameters

Answer: B

Q87. What is a test set used for?

A. Training the model on labeled training samples
B. Tuning the model's hyperparameter settings
C. Cleaning and preprocessing the raw data
D. Evaluating the model's performance on unseen data

Answer: D

Q88. What is a false positive?

A. Making a correct and accurate model prediction
B. Predicting positive when the actual value is negative
C. Encountering a missing value in the raw data
D. Predicting negative when the actual value is positive

Answer: B

Q89. R-squared measures:

A. The proportion of variance in the target explained by the model
B. The total number of data samples in the dataset
C. The total number of input features in the training dataset
D. The total elapsed time required for model training

Answer: A

Q90. What is a validation set?

A. The same set as the independent held-out test set
B. The entire full dataset including all partitions
C. A preprocessing data cleaning utility pipeline
D. A subset used to tune hyperparameters during training

Answer: D

Q91. What is feature engineering?

A. Building custom computer hardware for faster processing
B. Creating new features from existing data to improve model performance
C. Deleting all features from the dataset before model training
D. Only using the original raw data without any transformations

Answer: B

Q92. What is a feature?

A. A specific type of learning algorithm
B. A particular model training method
C. An input variable used by the model for prediction
D. The output produced by a trained model

Answer: A

Q93. What is feature selection?

A. Training the model on the full dataset
B. Choosing the most relevant features for the model
C. Deleting the entire dataset from storage
D. Creating new derived features from raw data

Answer: B

Q94. What is a binary feature?

A. A continuous numeric feature
B. A missing feature value
C. A feature with many distinct categories
D. A feature with only two possible values (0 or 1)

Answer: C

Q95. What is a numerical feature?

A. A feature with categorical text categories
B. A feature containing image data
C. A feature containing audio data
D. A feature with continuous or discrete numeric values

Answer: A

Q96. Feature extraction involves:

A. Deleting features from a dataset
B. Deriving new features from raw data
C. Renaming features in a table
D. Copying features between datasets

Answer: B

Q97. What does dropping a feature mean?

A. Removing a feature from the dataset
B. Transforming a feature with a log
C. Adding a brand new feature column
D. Scaling a feature to unit range

Answer: A

Q98. Why might you create interaction features?

A. To remove extreme outlier values from data
B. To normalize data to a standard range
C. To capture relationships between two or more features
D. To reduce the overall size of the dataset

Answer: C

Q99. What is binning?

A. Removing selected features from the dataset
B. Adding new synthetic features to the data
C. Converting continuous features into discrete intervals
D. Sorting data records by a column value

Answer: C

Q100. What is a datetime feature?

A. A feature representing dates and times
B. A purely categorical feature only
C. A strictly binary feature only
D. A purely numerical feature only

Answer: B

Q101. What is a neural network?

A. A network routing protocol for data packets
B. A computing system inspired by biological neural networks
C. A comparison-based sorting algorithm for arrays
D. A relational database for storing structured records

Answer: B

Q102. What is a neuron (node) in a neural network?

A. A basic computational unit that receives inputs and produces an output
B. A physical data storage unit for persisting files on disk
C. A specific type of input feature used in the dataset
D. A gradient-based training algorithm for optimization

Answer: A

Q103. What is an activation function?

A. A technique for initializing random weight parameters
B. A function that introduces non-linearity into the network
C. A function for loading data from external file sources
D. A method for calculating the total loss of the model

Answer: B

Q104. What is a hidden layer?

A. A data preprocessing transformation layer
B. The very last output layer of the network
C. The very first input layer of the network
D. A layer between the input and output layers

Answer: D

Q105. What is the ReLU activation function?

A. f(x) = x^2
B. f(x) = 1/(1+e^-x)
C. f(x) = max(0, x)
D. f(x) = tanh(x)

Answer: C

Q106. What is backpropagation?

A. A type of non-linear activation function for neurons
B. A specific deep neural network architecture design
C. A data preprocessing technique for cleaning raw inputs
D. An algorithm for computing gradients to update network weights

Answer: D

Q107. Which framework is commonly used for deep learning?

A. NumPy only
B. Flask
C. TensorFlow
D. Django

Answer: C

Q108. What is an epoch in training?

A. A type of non-linear activation function node
B. A single individual data point in the training set
C. A single hidden layer within the neural network
D. One complete pass through the entire training dataset

Answer: D

Q109. What is a loss function?

A. An individual input feature column used for training
B. A non-linear activation function applied at hidden layers
C. A structured collection of labeled data for evaluation
D. A function measuring how well predictions match actual values

Answer: D

Q110. What is gradient descent?

A. A type of multi-layer feedforward neural network architecture with connections
B. An optimization algorithm minimizing loss by updating weights via steepest descent
C. A tree-based hierarchical data structure for nearest-neighbor lookups
D. A non-linear activation function that maps inputs to the range zero to one

Answer: B

Q111. What is a Convolutional Neural Network (CNN)?

A. A neural network designed for processing grid-like data such as images
B. A relational database system for storing structured records
C. A recurrent network designed exclusively for text sequence data
D. A comparison-based sorting algorithm for numerical arrays

Answer: A

Q112. What is a Recurrent Neural Network (RNN)?

A. A neural network for sequential data with memory of previous inputs
B. A convolutional network designed exclusively for image pixel data
C. An unsupervised density-based clustering algorithm for grouping
D. A static feedforward network with no temporal connections

Answer: A

Q113. What is a GAN?

A. An unsupervised clustering method based on density estimation
B. A specialized variant of convolutional neural networks for images
C. Generative Adversarial Network - two networks competing to generate realistic data
D. A supervised regression algorithm for predicting numeric values

Answer: C

Q114. What is an autoencoder?

A. A neural network that learns to compress and reconstruct its input
B. A comparison-based sorting algorithm for numerical arrays
C. A data loading utility for reading files into memory
D. A supervised classifier that predicts categorical labels from data

Answer: A

Q115. What is LSTM?

A. Long Short-Term Memory - an RNN variant handling long-term dependencies
B. A cross-entropy loss function for multi-class classification
C. A specialized type of convolutional neural network for images
D. A structured data format for storing tabular information

Answer: A

Q116. What is a pooling layer in a CNN?

A. A data loading layer for reading external files
B. The final output layer for class predictions
C. A layer that increases the total number of neurons
D. A layer that reduces spatial dimensions by down-sampling

Answer: D

Q117. What is the purpose of a convolutional filter (kernel)?

A. To add random noise for data augmentation purposes
B. To store training data on disk for later retrieval access
C. To increase the spatial resolution of the input image
D. To detect specific features like edges or textures in input data

Answer: D

Q118. What is a pre-trained model?

A. A randomly initialized model with no learned representations
B. A model with absolutely no prior training on any data
C. A model that has been permanently deleted from storage
D. A model already trained on a large dataset for use on new tasks

Answer: D

Q119. What is a generative model?

A. A model that only performs supervised categorical classification
B. A model that only performs numerical regression prediction
C. A model that learns to generate new data similar to training data
D. A model that only performs unsupervised data point clustering

Answer: C

Q120. What framework was developed by Facebook (Meta) for deep learning?

A. Caffe
B. Keras
C. PyTorch
D. TensorFlow

Answer: C

Q121. What does NLP stand for?

A. Numerical Linear Processing
B. Network Logic Protocol
C. Natural Language Processing
D. Neural Language Programming

Answer: C

Q122. Tokenization in NLP means:

A. Encrypting text for secure data transmission
B. Splitting text into individual words or subwords
C. Deleting text from the data pipeline
D. Compressing text to reduce storage size

Answer: B

Q123. What is sentiment analysis?

A. Summarizing long documents into shorter text passages
B. Translating text from one natural language to another language
C. Generating entirely new text from a given prompt input
D. Determining the emotional tone of text (positive, negative, neutral)

Answer: D

Q124. Stop words are:

A. Important domain-specific keywords carrying high information
B. Common words like 'the' and 'is' often removed in processing
C. Rare words that appear only once in the entire corpus
D. Technical terms from a specialized subject vocabulary

Answer: B

Q125. What is text classification?

A. Generating entirely new text from a prompt
B. Translating text between different languages
C. Editing and correcting text for grammar
D. Assigning predefined categories to text documents

Answer: D

Q126. What is a corpus in NLP?

A. A large collection of text documents
B. A specific grammar rule definition
C. A single word in a sentence
D. A type of neural network model

Answer: A

Q127. What is stemming?

A. Reducing words to their root form by removing suffixes
B. Adding prefixes to words for morphological expansion
C. Counting the total frequency of words in text
D. Translating words between two different languages

Answer: A

Q128. What is lemmatization?

A. Counting the total characters in a word
B. Removing all vowels from the input text
C. Adding suffixes for morphological inflection
D. Reducing words to their base dictionary form

Answer: D

Q129. Named Entity Recognition (NER) identifies:

A. Grammar errors found throughout the given text
B. Spelling mistakes identified within the text body
C. The overall length of each sentence in the document
D. Named entities like people, organizations, and locations in text

Answer: A

Q130. What is machine translation?

A. Performing manual text translation
B. Generating new text from a prompt
C. Summarizing text into shorter form
D. Automatically translating text from one language to another

Answer: C

Q131. Computer vision enables computers to:

A. Interpret and understand visual information from images and videos
B. Design and build responsive web page user interfaces
C. Manage and query relational database table records
D. Process only natural language text data and documents

Answer: A

Q132. What is image classification?

A. Segmenting an image into parts
B. Assigning a label to an entire image
C. Detecting objects in a given image
D. Generating an image from scratch

Answer: C

Q133. What is a pixel?

A. An input feature column name
B. A gradient training algorithm
C. A type of deep neural network
D. The smallest unit of a digital image

Answer: D

Q134. What is object detection?

A. Removing detected objects from an existing image frame
B. Classifying the entire image into a single category label
C. Identifying and locating objects within an image with bounding boxes
D. Generating entirely new photorealistic synthetic images

Answer: C

Q135. RGB stands for:

A. Random, Generated, Binary
B. Recursive, Generative, Base
C. Real, Gradient, Bias
D. Red, Green, Blue

Answer: D

Q136. What is image segmentation?

A. Resizing an image to different specified dimensions
B. Partitioning an image into meaningful regions at the pixel level
C. Rotating an image by a specified number of degrees
D. Cropping an image to a smaller rectangular area

Answer: C

Q137. What does a convolution operation do to an image?

A. Applies a filter to extract features like edges and textures
B. Increases the image resolution to a higher pixel count
C. Changes the file format of the image container file
D. Deletes the image entirely from disk storage space

Answer: A

Q138. What is a grayscale image?

A. An image with only shades of gray (single channel)
B. A volumetric three-dimensional image stack
C. A binary black and white image with two values
D. A full-color image with three RGB channels

Answer: A

Q139. What is image resizing?

A. Rotating the image by a specified angle value
B. Changing the dimensions (width and height) of an image
C. Deleting the image from the disk file system
D. Changing the color space of the image channels

Answer: B

Q140. Face detection is an example of:

A. Reinforcement learning
B. Computer vision
C. Data analytics
D. Natural language processing

Answer: B

Q141. Big Data is characterized by:

A. Only the fast speed of arrival
B. Only the diverse data types
C. Volume, Velocity, Variety (3 Vs)
D. Only the large volume of data

Answer: C

Q142. What is Hadoop?

A. A standalone relational database for transactional queries
B. A multi-layer deep neural network architecture for AI
C. An open-source framework for distributed storage and processing of big data
D. A general-purpose compiled programming language for applications

Answer: C

Q143. What is MapReduce?

A. A comparison-based sorting algorithm for small arrays
B. A programming model for processing large datasets in parallel across a cluster
C. A structured query language for relational database management
D. A supervised machine learning algorithm for classification

Answer: B

Q144. What is Apache Spark?

A. A production web server for hosting applications
B. A standalone relational database management system
C. A general-purpose programming language runtime
D. A fast distributed computing engine for big data processing

Answer: D

Q145. What is a data warehouse?

A. A centralized repository for structured data used for analysis and reporting
B. A multi-layer deep neural network trained on data
C. A physical room for storing server hardware equipment
D. A transactional database used for daily operational queries

Answer: A

Q146. What does ETL stand for?

A. Edit, Transfer, Link
B. Extract, Transform, Load
C. Encrypt, Transmit, Log
D. Evaluate, Test, Learn

Answer: B

Q147. What is a data lake?

A. A cleaned and normalized relational database table store
B. An interactive dashboard for data visualization tools
C. A storage repository holding raw data in its native format
D. A small curated dataset for specific model training

Answer: C

Q148. What is distributed computing?

A. Cloud storage of files without any computation tasks
B. Spreading computation across multiple machines working together
C. Using a single powerful dedicated computer for processing
D. Interactive data visualization and charting tools

Answer: B

Q149. What is HDFS?

A. A general-purpose interpreted programming language runtime
B. A desktop operating system for personal computers
C. A multi-layer deep neural network for classification
D. Hadoop Distributed File System for storing data across a cluster

Answer: D

Q150. What is batch processing?

A. Processing data interactively with user input prompts
B. Processing data manually without any automation
C. Processing large volumes of data as a group at scheduled intervals
D. Processing data in real-time as each record arrives

Answer: C

Q151. What is MLOps?

A. Only the model training phase of the ML lifecycle
B. Only the data visualization phase of the lifecycle
C. Practices for deploying and maintaining ML models in production
D. Only the data collection phase of the ML lifecycle

Answer: C

Q152. What is model deployment?

A. Collecting and gathering raw data from sources
B. Making a trained model available for use in production
C. Training a model on labeled training data samples
D. Preprocessing and cleaning the raw input data

Answer: B

Q153. What is an API in the context of model deployment?

A. A gradient-based training algorithm for optimizing model weights
B. An interactive visualization tool for exploring model results
C. An interface allowing applications to request predictions from the model
D. A relational database for storing structured training data

Answer: C

Q154. What is model versioning?

A. Tracking different versions of a model over time
B. Deleting old models from the storage system
C. Reducing model size through weight pruning
D. Training models to converge significantly faster

Answer: A

Q155. What is a Docker container?

A. A physical rack-mounted server housed in a data center
B. A standalone relational database for transactional queries
C. A multi-layer deep neural network model architecture
D. A lightweight portable package containing everything to run an application

Answer: D

Q156. What is CI/CD in MLOps?

A. An interactive data visualization and dashboard charting tool
B. A standalone relational database management query system
C. A general-purpose compiled programming language for building apps
D. Continuous Integration and Continuous Deployment for automating ML pipelines

Answer: D

Q157. What is model monitoring?

A. Collecting and gathering new raw data from external sources
B. Training new models from scratch on fresh training data
C. Building automated data preprocessing pipeline workflows
D. Tracking model performance and behavior in production over time

Answer: D

Q158. What is a REST API?

A. A gradient-based model training optimization method
B. A type of relational database storage engine format
C. An API architecture using HTTP methods for communication
D. A general-purpose compiled programming language

Answer: C

Q159. What does reproducibility mean in ML?

A. Using less training data to save storage and costs
B. Running trained models faster by using better hardware
C. The ability to recreate the same results with the same data and code
D. Making models larger by adding more network layers

Answer: C

Q160. What is a model artifact?

A. The saved output of training including model weights and configuration
B. Only the source code used to build and train the model
C. Only the runtime logs generated during model training
D. Only the raw training data used during model training

Answer: A

Q161. What is AI bias?

A. Systematic unfairness in AI predictions due to biased data or algorithms
B. Slow processing speed during model inference time
C. Random errors that occur during the model prediction process
D. Hardware failures that cause incorrect output computations

Answer: A

Q162. What is data privacy?

A. Encrypting only the physical hardware server equipment
B. Making all collected data publicly available and open
C. Deleting all collected data from every storage system
D. Protecting personal information from unauthorized access and use

Answer: D

Q163. What is informed consent in AI?

A. Sharing collected data freely with any third party entity
B. Ignoring user preferences about their data usage rights
C. Ensuring people understand and agree to how their data will be used
D. Collecting personal data without any prior user permission

Answer: C

Q164. What is transparency in AI?

A. Making AI decision-making processes understandable and open
B. Making AI systems run faster and more efficiently
C. Hiding how AI systems work from all stakeholders
D. Using significantly more data for model training

Answer: A

Q165. What is the GDPR?

A. A European regulation for data protection and privacy
B. A deep neural network architecture for images
C. A supervised machine learning training algorithm
D. A general-purpose compiled programming language

Answer: A

Q166. Why is fairness important in AI?

A. To use less training data during the model building phase
B. To ensure AI systems treat all groups equitably without discrimination
C. To make AI systems process data faster and more efficiently
D. To reduce the financial costs of developing AI systems

Answer: B

Q167. What is accountability in AI?

A. Automating every aspect without any human involvement
B. Ensuring responsible parties can be identified when AI causes harm
C. Making AI systems completely autonomous without oversight
D. Removing all human oversight from AI decision processes

Answer: B

Q168. What is an adversarial attack on AI?

A. Deliberately crafted inputs designed to fool AI systems
B. A physical attack on the server hardware equipment
C. A network attack on the communication infrastructure
D. A SQL injection attack on the database system

Answer: A

Q169. What is algorithmic fairness?

A. Making algorithms process data faster through hardware optimization
B. Reducing the overall storage size of trained model files
C. Using more diverse and larger training datasets for models
D. Ensuring algorithms make decisions without discrimination on protected attributes

Answer: D

Q170. What is the purpose of AI regulations?

A. To ensure safe fair and responsible development and use of AI
B. To make AI systems significantly more expensive to build
C. To deliberately slow down all AI research and development
D. To limit all AI capabilities to only narrow simple tasks

Answer: A

Q171. What type of AI is designed to perform a single specific task?

A. Meta AI
B. Narrow AI
C. General AI
D. Super AI

Answer: B

Q172. Which term describes a machine's ability to imitate human conversation?

A. Chatbot
B. Compiler
C. Web Scraping
D. Data Mining

Answer: A

Q173. What does the term 'training data' refer to in machine learning?

A. Data transmitted between two network endpoint nodes
B. Data stored in a database for backup recovery use
C. Data used to teach a model to recognize patterns
D. Data used to evaluate final model accuracy scores

Answer: C

Q174. Which field combines statistics, algorithms, and computing to extract insights?

A. Data Science
B. Web Design
C. Compiling
D. Networking

Answer: A

Q175. What is reinforcement learning primarily based on?

A. Clustering of similar items
B. Labeled training datasets
C. Manual rule programming
D. Reward and penalty signals

Answer: D

Q176. Which of the following best describes deep learning?

A. Learning restricted to tabular structured datasets
B. Learning without any computational hardware need
C. Learning from only a single data sample at once
D. Learning using neural networks with many layers

Answer: D

Q177. What is the Turing Test designed to evaluate?

A. A machine's intelligent behavior
B. Memory capacity of a system
C. Energy efficiency of a chip
D. Processing speed of hardware

Answer: A

Q178. Which AI application converts spoken language into written text?

A. Data Warehousing
B. Image Segmentation
C. Speech Recognition
D. Load Balancing

Answer: C

Q179. What does the abbreviation ML stand for in the context of AI?

A. Machine Learning
B. Meta Language
C. Model Linking
D. Memory Loading

Answer: A

Q180. Which type of learning uses labeled examples to train a model?

A. Unsupervised learning
B. Supervised learning
C. Reinforcement learning
D. Transfer learning

Answer: B

Q181. What is a scalar in linear algebra?

A. A vector of values
B. A tensor of arrays
C. A matrix of numbers
D. A single numerical value

Answer: D

Q182. What does the dot product of two vectors produce?

A. A new tensor
B. A new matrix
C. A single scalar
D. A new vector

Answer: C

Q183. What is the mean of the numbers 2, 4, 6, 8, and 10?

A. 6
B. 7
C. 4
D. 5

Answer: A

Q184. Which operation flips a matrix over its diagonal?

A. Transpose
B. Rotation
C. Inversion
D. Reduction

Answer: A

Q185. What does a probability value of zero represent?

A. A random event
B. Impossible event
C. A likely event
D. A certain event

Answer: B

Q186. What is the derivative of a constant value?

A. Infinity
B. Undefined
C. One
D. Zero

Answer: D

Q187. Which term describes a rectangular array of numbers in rows and columns?

A. Scalar
B. Matrix
C. Vector
D. String

Answer: B

Q188. What is the range of a standard probability value?

A. Negative one to one
B. Minus ten to ten
C. Zero to one only
D. Zero to infinity

Answer: C

Q189. What does the summation symbol sigma represent?

A. Ratio of values
B. Product of terms
C. Limit of terms
D. Sum of a series

Answer: D

Q190. What is an identity matrix?

A. A matrix with all zeros in every position
B. A matrix with all elements equal to two
C. A square matrix with ones on diagonal only
D. A rectangular matrix with random entries

Answer: C

Q191. Which Python library is primarily used for numerical array computations?

A. Flask
B. NumPy
C. Django
D. Tkinter

Answer: B

Q192. What does the pandas function read_csv() do?

A. Compresses CSV into zip
B. Writes data to a CSV file
C. Reads CSV into a DataFrame
D. Deletes a CSV from disk

Answer: C

Q193. Which library is most commonly used for creating plots in Python?

A. Logging
B. Subprocess
C. Requests
D. Matplotlib

Answer: D

Q194. What data structure does pandas use to represent tabular data?

A. Binary Tree
B. DataFrame
C. Linked List
D. Dictionary

Answer: B

Q195. Which keyword is used to define a function in Python?

A. function
B. func
C. def
D. define

Answer: C

Q196. What does the shape attribute of a NumPy array return?

A. The dimensions of array
B. Total memory in bytes
C. The maximum array value
D. The data type of elements

Answer: A

Q197. Which Python library provides algorithms like SVM and Random Forest?

A. Beautiful Soup
B. SQLAlchemy
C. Pillow Library
D. Scikit-learn

Answer: D

Q198. What is a Python list comprehension used for?

A. Deleting items from dictionary
B. Creating lists with concise syntax
C. Handling file system exceptions
D. Connecting to remote databases

Answer: B

Q199. Which method checks for missing values in a pandas DataFrame?

A. dropna()
B. notna()
C. fillna()
D. isnull()

Answer: D

Q200. What does pip install do in Python?

A. Runs Python unit tests
B. Creates virtual environments
C. Installs Python packages
D. Compiles Python source code

Answer: C

Q201. What is data cleaning in the context of data preprocessing?

A. Compressing data for storage
B. Adding noise to the data
C. Encrypting data for security
D. Fixing errors and removing inconsistencies

Answer: D

Q202. Which technique fills missing values with the column average?

A. Normalization step
B. Mode imputation
C. Deletion method
D. Mean imputation

Answer: D

Q203. What does one-hot encoding convert?

A. Categories into binary vectors
B. Numbers into text labels
C. Images into pixel arrays
D. Audio into text transcripts

Answer: A

Q204. What is the purpose of removing duplicate records from a dataset?

A. To add more noise and improve model robustness
B. To reduce the number of features in the dataset
C. To avoid bias from repeated identical observations
D. To increase the total number of training samples

Answer: C

Q205. Which method scales features to a range between zero and one?

A. Standardization
B. Log transformation
C. Min-max normalization
D. Binning technique

Answer: C

Q206. What is a common file format for storing tabular datasets?

A. CSV
B. PNG
C. MP3
D. EXE

Answer: A

Q207. What does the term 'missing data' refer to in a dataset?

A. Data stored in a different file location
B. Values that are absent for some observations
C. Values that are larger than the average value
D. Data that has been encrypted for privacy

Answer: B

Q208. What is the first step typically performed in any data preprocessing pipeline?

A. Data collection
B. Hyperparameter tuning
C. Model training
D. Feature selection

Answer: A

Q209. Which type of data contains categories like color or country name?

A. Numerical data
B. Sequential data
C. Continuous data
D. Categorical data

Answer: D

Q210. What does the term 'outlier' refer to in a dataset?

A. A value significantly different from others
B. The average value of all observations
C. A missing entry in a data column
D. The most common value in the data

Answer: A

Q211. What does EDA stand for in data science?

A. Extended Data Algorithm
B. Extra Data Analysis
C. External Data Approach
D. Exploratory Data Analysis

Answer: D

Q212. Which type of chart is best for showing the distribution of a single variable?

A. Pie chart
B. Line chart
C. Histogram
D. Gantt chart

Answer: C

Q213. What does a scatter plot visualize?

A. The hierarchical structure of a dataset
B. The proportion of categories in a dataset
C. The trend of a variable over time periods
D. The relationship between two numerical variables

Answer: D

Q214. What is the median of a sorted dataset?

A. The middle value in order
B. The range of the data
C. The arithmetic average
D. The most frequent value

Answer: A

Q215. Which plot type shows the five-number summary of a dataset?

A. Bar chart
B. Area chart
C. Box plot
D. Scatter plot

Answer: C

Q216. What does a correlation coefficient of +1 indicate?

A. No relationship at all
B. Perfect positive linear
C. Random relationship
D. Perfect negative linear

Answer: B

Q217. Which visualization is best for comparing quantities across categories?

A. Line chart
B. Bar chart
C. Heat map
D. Histogram

Answer: B

Q218. What is the mode of a dataset?

A. The most frequent value
B. The largest value
C. The average value
D. The middle value

Answer: A

Q219. What does a line chart typically display?

A. Trends over time
B. Data distribution
C. Category proportions
D. Feature correlations

Answer: A

Q220. What is the purpose of a heatmap in data analysis?

A. To compress images into smaller sizes
B. To display magnitude of values using colors
C. To show geographic locations on a map
D. To play audio data as visual patterns

Answer: B

Q221. What is the goal of a supervised learning algorithm?

A. Generate new data from scratch
B. Learn a mapping from inputs to outputs
C. Find hidden patterns without labels
D. Group similar data points together

Answer: B

Q222. Which of these is a classification algorithm?

A. Principal Component
B. Linear Regression
C. K-Means Clustering
D. Logistic Regression

Answer: D

Q223. What does a decision tree use to split data at each node?

A. Feature thresholds
B. Random selection
C. Output labels only
D. Time stamps alone

Answer: A

Q224. What type of problem predicts a continuous numerical value?

A. Clustering task
B. Association task
C. Classification task
D. Regression task

Answer: D

Q225. Which algorithm finds the best-fit line through data points?

A. Random Forest model
B. K-Nearest Neighbors
C. Apriori Algorithm
D. Linear Regression

Answer: D

Q226. What does KNN stand for in machine learning?

A. K-Null Notation
B. K-Nearest Neighbors
C. K-Node Navigation
D. K-Normal Networks

Answer: B

Q227. In binary classification, how many possible output classes exist?

A. Three
B. Two
C. One
D. Four

Answer: B

Q228. What is a label in supervised learning?

A. An input feature column
B. The correct output answer
C. A training epoch count
D. A model parameter value

Answer: B

Q229. Which algorithm is often used for predicting house prices?

A. Linear Regression model
B. Naive Bayes classifier
C. K-Means algorithm
D. Apriori association rule

Answer: A

Q230. What does the term 'training' mean in supervised learning?

A. Visualizing data in chart format
B. Manually coding rules for predictions
C. Adjusting model parameters using data
D. Removing features from the dataset

Answer: C

Q231. What is an ensemble method in machine learning?

A. Combining multiple models
B. A data cleaning technique
C. A single complex model
D. A visualization approach

Answer: A

Q232. Which ensemble method trains models on random data subsets?

A. Stacking
B. Bagging
C. Pruning
D. Boosting

Answer: B

Q233. What is Random Forest built upon?

A. Multiple logistic regression models combined together
B. Multiple SVM models with different kernel functions
C. Multiple neural networks of varying layer depths
D. Multiple decision trees with random feature subsets

Answer: D

Q234. How does a majority voting ensemble make predictions?

A. Using the most common prediction
B. Using the maximum prediction
C. Using the average prediction value
D. Using the minimum prediction

Answer: A

Q235. What is the main advantage of ensemble methods over single models?

A. They need less data input
B. They always train faster
C. They use less memory
D. They reduce overall error

Answer: D

Q236. Which technique trains models sequentially to correct prior errors?

A. Bagging
B. Boosting
C. Blending
D. Stacking

Answer: B

Q237. What does the term 'base learner' mean in ensemble learning?

A. The evaluation metric used
B. An individual model in ensemble
C. The training data subset used
D. The final combined model

Answer: B

Q238. Which of these is a popular gradient boosting library?

A. XGBoost
B. NumPy
C. Matplotlib
D. Pandas

Answer: A

Q239. What is the purpose of bootstrap sampling in bagging?

A. To select the best features
B. To normalize feature values
C. To create diverse training sets
D. To remove outliers from data

Answer: C

Q240. How does Random Forest handle the prediction for classification tasks?

A. Uses the prediction of the last tree only
B. Uses the prediction of the first tree only
C. Takes majority vote from all the trees
D. Averages all tree predictions numerically

Answer: C

Q241. What is unsupervised learning?

A. Learning from labeled data
B. Learning with human feedback
C. Learning from test data only
D. Learning without any labels

Answer: D

Q242. What is the goal of clustering in unsupervised learning?

A. Predicting a target value
B. Generating synthetic data
C. Grouping similar data points
D. Reducing training time

Answer: C

Q243. Which algorithm groups data into K predefined number of clusters?

A. K-Nearest Neighbors
B. K-Fold Validation
C. K-Means Clustering
D. K-Mode Analysis

Answer: C

Q244. What is dimensionality reduction used for?

A. Increasing the sample count
B. Reducing the number of features
C. Converting data to text format
D. Adding more features to data

Answer: B

Q245. Which technique is a common dimensionality reduction method?

A. Principal Component Analysis
B. Logistic Regression
C. Random Forest model
D. Gradient Boosting

Answer: A

Q246. What is the centroid in K-Means clustering?

A. The smallest value in a cluster
B. The label assigned to a cluster
C. The data point farthest from cluster
D. The center point of a cluster

Answer: D

Q247. Which type of learning is used for customer segmentation?

A. Supervised learning
B. Reinforcement learning
C. Unsupervised learning
D. Semi-supervised learning

Answer: C

Q248. What does anomaly detection aim to find?

A. The most common patterns
B. The best classification model
C. Unusual or rare data points
D. The optimal feature weights

Answer: C

Q249. What is a dendrogram used to visualize?

A. Neural network architecture
B. Linear regression coefficients
C. Hierarchical clustering results
D. Confusion matrix of a model

Answer: C

Q250. Which algorithm does not require specifying the number of clusters beforehand?

A. Mini-Batch K-Means
B. K-Medoids
C. K-Means
D. DBSCAN

Answer: D

Q251. What is a confusion matrix used for?

A. Visualizing data distributions
B. Evaluating classification results
C. Preprocessing raw data
D. Training neural networks

Answer: B

Q252. What does accuracy measure in classification?

A. The number of features used
B. The proportion of correct predictions
C. The size of the training data
D. The speed of model training

Answer: B

Q253. What is the purpose of splitting data into training and test sets?

A. To remove outliers from the original dataset
B. To increase the training time of the model
C. To reduce the total amount of data available
D. To evaluate how well the model generalizes

Answer: D

Q254. What does precision measure in a classification model?

A. The overall accuracy of the model on all classes
B. The speed of prediction for each data sample
C. The proportion of actual positives correctly identified
D. The proportion of positive predictions that are correct

Answer: D

Q255. What does recall measure in a classification model?

A. The proportion of actual positives correctly identified
B. The number of features used by the trained model
C. The computational time required for each prediction
D. The proportion of positive predictions that are correct

Answer: A

Q256. What is cross-validation used for?

A. Generating synthetic data
B. Reducing the number of features
C. Assessing model performance robustly
D. Training the final production model

Answer: C

Q257. What is overfitting in the context of model evaluation?

A. The model performs well on training but poorly on test data
B. The model performs poorly on training data
C. The model has too few parameters
D. The model takes too long to train

Answer: A

Q258. What does the F1 score combine?

A. Bias and variance
B. Training and test error
C. Precision and recall
D. Accuracy and speed

Answer: C

Q259. What is a validation set used for?

A. Testing the final model performance
B. Training the model parameters
C. Cleaning and preprocessing the data
D. Tuning hyperparameters during development

Answer: D

Q260. What does underfitting indicate about a model?

A. The model is too simple to capture patterns
B. The model uses too many features
C. The model has been trained too long
D. The model is too complex for the data

Answer: A

Q261. What is feature engineering in machine learning?

A. Evaluating model performance
B. Training a model automatically
C. Creating informative input variables
D. Deploying models to production

Answer: C

Q262. What is feature selection?

A. Adding all available features to the model without any filtering
B. Converting feature names from one language to another one
C. Generating entirely new features from random noise data
D. Choosing the most relevant features for model training use

Answer: D

Q263. What does one-hot encoding create from a categorical feature?

A. A single numerical column
B. A text description field
C. Multiple binary columns
D. A sorted ordinal value

Answer: C

Q264. What is the purpose of scaling numerical features?

A. To add missing value indicators
B. To bring features to a similar range
C. To change the data type of features
D. To remove features from the data

Answer: B

Q265. Which technique creates new features by combining existing ones?

A. Feature deletion
B. Feature encryption
C. Feature interaction
D. Feature imputation

Answer: C

Q266. What is binning used for in feature engineering?

A. Encrypting sensitive data for privacy
B. Converting categories into continuous values
C. Removing duplicate values from dataset
D. Converting continuous variables into categories

Answer: D

Q267. What does a polynomial feature transformation create?

A. Compressed versions of original data
B. Lower dimensional representations
C. Higher order combinations of features
D. Random noise features for testing

Answer: C

Q268. What is the difference between ordinal and nominal encoding?

A. Ordinal preserves category order while nominal treats categories as equal
B. Ordinal only works with numbers while nominal only works with strings
C. Both encodings are identical and always produce the same output values
D. Ordinal treats categories as equal while nominal preserves their ordering

Answer: A

Q269. What is a derived feature?

A. An original feature from raw data
B. A new feature created from existing ones
C. A feature with many missing values
D. A feature removed during cleaning

Answer: B

Q270. Why is feature engineering important for model performance?

A. It always makes models train faster
B. It removes the requirement for training data
C. It eliminates the need for model tuning
D. Good features can significantly improve accuracy

Answer: D

Q271. What is a neural network inspired by?

A. Database systems
B. Computer circuits
C. The human brain
D. Mathematical proofs

Answer: C

Q272. What is an activation function in a neural network?

A. A function that loads training data
B. A function that saves the model
C. A function that introduces non-linearity
D. A function that cleans input data

Answer: C

Q273. What is a hidden layer in a neural network?

A. A layer between input and output
B. The output layer of the network
C. The first layer receiving raw data
D. The layer that stores data on disk

Answer: A

Q274. Which activation function outputs values between 0 and 1?

A. Linear function
B. Tanh function
C. ReLU function
D. Sigmoid function

Answer: D

Q275. What does the term 'epoch' mean in deep learning training?

A. The initialization of model weight values
B. One complete pass through all training data
C. A single forward pass of one sample
D. The final test evaluation of the model

Answer: B

Q276. What is the purpose of backpropagation in neural networks?

A. To calculate and propagate error gradients
B. To make forward predictions on data
C. To load training data into memory
D. To visualize the network architecture

Answer: A

Q277. What is a weight in a neural network?

A. The number of layers in the network
B. The learning rate of the optimizer
C. A learnable parameter connecting neurons
D. The size of the training dataset

Answer: C

Q278. What is the most commonly used activation function in hidden layers today?

A. Sigmoid
B. Softmax
C. Linear
D. ReLU

Answer: D

Q279. What does the output layer of a classification network produce?

A. Hidden representations
B. Class probability scores
C. Gradient computations
D. Raw pixel values

Answer: B

Q280. What is a bias term in a neural network?

A. The rate at which the model learns from training data
B. The preference of the model for certain training samples
C. The error in predictions caused by insufficient data size
D. An additional learnable parameter added to weighted sums

Answer: D

Q281. What does CNN stand for in deep learning?

A. Central Neural Network
B. Circular Neuron Network
C. Convolutional Neural Network
D. Connected Node Network

Answer: C

Q282. What does RNN stand for in deep learning?

A. Recursive Node Network
B. Recurrent Neural Network
C. Regional Neuron Network
D. Random Neural Network

Answer: B

Q283. What is a convolutional filter used for in CNNs?

A. Storing training labels
B. Detecting patterns in input
C. Compressing the model size
D. Removing noise from data

Answer: B

Q284. What is the purpose of a pooling layer in a CNN?

A. To add more features to data
B. To increase image resolution
C. To reduce spatial dimensions
D. To convert images to text

Answer: C

Q285. What type of data are RNNs primarily designed to handle?

A. Tabular data
B. Image data
C. Graph data
D. Sequential data

Answer: D

Q286. What is a GAN in deep learning?

A. Grouped Attention Network
B. Gradient Adjusted Network
C. General Analysis Network
D. Generative Adversarial Network

Answer: D

Q287. What is max pooling in a CNN?

A. Taking the average value from a region
B. Taking the median value from a region
C. Taking the maximum value from a region
D. Taking the minimum value from a region

Answer: C

Q288. What is an autoencoder used for?

A. Generating adversarial examples
B. Learning compressed data representations
C. Tuning model hyperparameters
D. Supervised classification tasks

Answer: B

Q289. What does LSTM stand for?

A. Layered Spatial Tensor Model
B. Local Statistical Testing Method
C. Long Short-Term Memory
D. Linear Sequential Training Model

Answer: C

Q290. What is the Transformer architecture primarily known for?

A. Reinforcement learning games
B. Image classification tasks
C. Self-attention mechanism for sequences
D. Database query optimization

Answer: C

Q291. What does NLP stand for?

A. Neural Logic Processing
B. Numerical Linear Programming
C. Natural Language Processing
D. Network Layer Protocol

Answer: C

Q292. What is tokenization in NLP?

A. Encrypting text for security use
B. Compressing text to save space
C. Splitting text into smaller units
D. Translating text between languages

Answer: C

Q293. What is sentiment analysis used for?

A. Detecting the emotion or opinion in text
B. Counting the words in a document
C. Correcting grammar errors in writing
D. Translating text to other languages

Answer: A

Q294. What is a stop word in natural language processing?

A. A common word often removed in preprocessing
B. A word that stops program execution
C. A word that contains a typing error
D. A word used only in technical documents

Answer: A

Q295. What is named entity recognition?

A. Identifying names of programming functions in code
B. Identifying proper nouns like people and places in text
C. Identifying the most frequent word in a paragraph text
D. Identifying syntax errors in written documents for fixes

Answer: B

Q296. What is stemming in text preprocessing?

A. Reducing words to their root or base form
B. Sorting words in alphabetical sequence
C. Adding prefixes to expand word meanings
D. Translating words to another language

Answer: A

Q297. What is a corpus in NLP?

A. A neural network architecture
B. A large collection of text data
C. A visualization chart format
D. A machine learning algorithm type

Answer: B

Q298. What does a language model predict?

A. The next word in a sequence
B. Hardware specifications
C. Database query results
D. Image classification labels

Answer: A

Q299. What is text classification in NLP?

A. Translating text between two languages
B. Converting speech to written text
C. Generating images from descriptions
D. Assigning categories to text documents

Answer: D

Q300. What is a word embedding?

A. A word encrypted for secure storage
B. A word stored in a database table
C. A word highlighted in a text editor
D. A dense vector representation of a word

Answer: D

Q301. What is computer vision?

A. A computer's screen resolution settings
B. AI field enabling machines to interpret images
C. A method for compressing video file data
D. A type of display monitor hardware

Answer: B

Q302. What is image classification?

A. Converting images to text format
B. Sorting images by file size on disk
C. Assigning a label to an entire image
D. Splitting an image into two halves

Answer: C

Q303. What does object detection identify in an image?

A. The resolution of the image file
B. The file format of the image data
C. The color profile of the image data
D. Objects and their bounding box locations

Answer: D

Q304. What is a pixel in digital image processing?

A. The smallest unit of a digital image
B. A color space conversion algorithm
C. A type of image compression format
D. A filter used to sharpen images

Answer: A

Q305. What does image segmentation do?

A. Converts color images to grayscale
B. Increases image resolution quality
C. Divides image into meaningful regions
D. Reduces the file size of images

Answer: C

Q306. What is data augmentation in computer vision?

A. Reducing the resolution of all images
B. Converting all images to black and white
C. Deleting duplicate images from dataset
D. Creating modified copies of training images

Answer: D

Q307. Which deep learning architecture is most associated with image processing?

A. Long Short-Term Memory
B. Generative Adversarial Network
C. Recurrent Neural Network
D. Convolutional Neural Network

Answer: D

Q308. What is face recognition technology used for?

A. Converting faces into cartoon drawings
B. Identifying or verifying a person's identity
C. Compressing facial images to save space
D. Measuring the resolution of face images

Answer: B

Q309. What is optical character recognition?

A. Recognizing and extracting text from images
B. Compressing text before saving to disk
C. Converting text files to images
D. Encrypting text within image files

Answer: A

Q310. What does RGB stand for in digital images?

A. Random Generated Bits
B. Red Green Blue
C. Regular Grid Base
D. Rapid Graphics Buffer

Answer: B

Q311. What are the three Vs of big data?

A. Vector, Variable, Vault
B. Version, Vendor, Virtue
C. Vision, Value, Validity
D. Volume, Velocity, Variety

Answer: D

Q312. What is Apache Hadoop used for?

A. Building web applications
B. Distributed storage and processing
C. Designing user interfaces
D. Creating mobile applications

Answer: B

Q313. What is Apache Spark?

A. A database management system
B. A fast distributed computing engine
C. A web development framework
D. A machine learning algorithm

Answer: B

Q314. What does HDFS stand for in big data?

A. High Data File System
B. Hardware Distribution Framework
C. Hybrid Data Format Standard
D. Hadoop Distributed File System

Answer: D

Q315. What is a data lake?

A. A small SQL database
B. A machine learning algorithm
C. A centralized repository for raw data
D. A data visualization tool

Answer: C

Q316. What is MapReduce in big data processing?

A. A database query language
B. A machine learning framework
C. A parallel processing programming model
D. A data visualization library

Answer: C

Q317. What is streaming data?

A. Data generated and processed continuously
B. Data archived for long-term storage
C. Data stored in static files
D. Data compressed for efficient transfer

Answer: A

Q318. What is a data warehouse?

A. An unstructured data lake for raw files
B. A tool for creating data visualizations
C. A physical storage building for servers
D. A structured repository for analyzed data

Answer: D

Q319. What is the purpose of data partitioning in big data systems?

A. To convert data from one format to another
B. To divide data across nodes for parallel processing
C. To encrypt data for security purposes
D. To delete old data from the storage system

Answer: B

Q320. What is Apache Kafka primarily used for?

A. Real-time data streaming and messaging
B. Image processing tasks
C. Creating web applications
D. Training deep learning models

Answer: A

Q321. What does MLOps stand for?

A. Machine Learning Operations
B. Multi-Layer Optimization
C. Machine Logic Operations
D. Model Linear Operations

Answer: A

Q322. What is model deployment in machine learning?

A. Training a model on data
B. Collecting new training data
C. Making a trained model available for use
D. Deleting an old model version

Answer: C

Q323. What is a REST API commonly used for in ML deployment?

A. Storing training data in databases
B. Serving model predictions over HTTP
C. Training models faster
D. Visualizing model architecture

Answer: B

Q324. What is model versioning?

A. Converting models between different formats
B. Deleting older models from storage
C. Tracking different versions of a trained model
D. Training multiple models simultaneously

Answer: C

Q325. What is CI/CD in the context of MLOps?

A. Calculated Input and Computed Derivation
B. Continuous Integration and Continuous Deployment
C. Computer Intelligence and Cloud Delivery
D. Central Interface and Core Distribution

Answer: B

Q326. What is model monitoring in production?

A. Training the model with new data
B. Deleting unused model files from disk
C. Tracking model performance after deployment
D. Creating documentation for the model

Answer: C

Q327. What is a Docker container used for in MLOps?

A. Storing large datasets for training
B. Visualizing model predictions on charts
C. Packaging applications with all dependencies
D. Training deep learning models faster

Answer: C

Q328. What is an ML pipeline?

A. A visualization of model accuracy metrics
B. A single model training step
C. A database for storing predictions
D. An automated sequence of ML workflow steps

Answer: D

Q329. What is A/B testing used for in model deployment?

A. Training two models at the same time
B. Splitting data into training and test sets
C. Merging two models into a single model
D. Comparing two model versions with real users

Answer: D

Q330. What is a feature store in MLOps?

A. A retail store that sells ML hardware
B. A centralized repository for ML features
C. A visualization tool for model features
D. A database for storing raw training data

Answer: B

Q331. What is AI bias?

A. A technique for improving model accuracy
B. A type of neural network optimization method
C. A method for collecting training data faster
D. Systematic unfairness in AI system outputs

Answer: D

Q332. What is data privacy in the context of AI?

A. Making all data publicly available online
B. Sharing data freely between all companies
C. Protecting personal information in AI systems
D. Storing data without any encryption method

Answer: C

Q333. What does explainable AI aim to provide?

A. Larger training datasets for models
B. More complex model architectures
C. Faster model training speeds
D. Understandable AI decision reasoning

Answer: D

Q334. What is fairness in AI systems?

A. Making AI models run on all hardware
B. Minimizing the size of the trained model
C. Ensuring equal treatment across different groups
D. Maximizing accuracy on the training dataset

Answer: C

Q335. What is the GDPR in relation to AI and data?

A. A European data protection regulation
B. A programming language for AI development
C. A type of neural network architecture
D. A machine learning algorithm for privacy

Answer: A

Q336. What is an adversarial attack on an AI system?

A. A method for training models faster now
B. A technique for compressing model size
C. Deliberately manipulating input to fool AI
D. Improving model accuracy by adding data

Answer: C

Q337. What does transparency mean in AI ethics?

A. Making AI code run more efficiently
B. Reducing the size of AI model files
C. Hiding model details from all users
D. Being open about how AI systems work

Answer: D

Q338. What is informed consent in AI data collection?

A. Automatically opting in all users to data sharing
B. Getting user agreement before collecting data
C. Collecting data without user knowledge
D. Deleting all collected data after processing

Answer: B

Q339. What is the purpose of AI regulation?

A. To prevent all AI research and development
B. To restrict AI use to government agencies only
C. To ensure AI systems are safe and responsible
D. To make AI technology more expensive to build

Answer: C

Q340. What is algorithmic transparency?

A. Making algorithms run faster on hardware
B. Using transparent display screens for output
C. Making AI decision processes understandable
D. Hiding algorithm details from competitors

Answer: C

Q341. Which branch of AI focuses on enabling machines to learn from data?

A. Robotics
B. Machine Learning
C. Computer Graphics
D. Database Management

Answer: B

Q342. What is a robot in the context of AI?

A. A software virus
B. A physical or virtual agent that can perform tasks autonomously
C. A type of database
D. A programming language

Answer: B

Q343. Which of the following is an application of natural language processing?

A. Image compression
B. Voice assistants like Siri
C. Video rendering
D. Battery management

Answer: B

Q344. What does the term 'intelligent agent' mean in AI?

A. A human expert
B. An entity that perceives its environment and takes actions to achieve goals
C. A computer virus
D. A data table

Answer: B

Q345. Which of the following is NOT an AI application?

A. Spam email filtering
B. Self-driving cars
C. Manual data entry on paper
D. Recommendation systems

Answer: C

Q346. What is computer vision in AI?

A. A way to improve monitor resolution
B. The field enabling computers to interpret visual information from the world
C. A type of encryption
D. A database query language

Answer: B

Q347. What type of data does a supervised learning algorithm require?

A. Only images
B. Labeled data with input-output pairs
C. Random noise
D. Encrypted data

Answer: B

Q348. What is an AI model?

A. A physical robot
B. A mathematical representation learned from data to make predictions
C. A type of hard drive
D. A monitor display

Answer: B

Q349. Which industry commonly uses AI for fraud detection?

A. Agriculture
B. Banking and Finance
C. Construction
D. Textile manufacturing

Answer: B

Q350. What is the purpose of a recommendation system?

A. To delete user accounts
B. To suggest relevant items to users based on their preferences
C. To compress files
D. To format documents

Answer: B

Q351. What is a vector in mathematics?

A. A single number
B. An ordered list of numbers with magnitude and direction
C. A type of graph
D. A random variable

Answer: B

Q352. What does the term 'dimension' refer to in a dataset?

A. The physical size of a hard drive
B. The number of features or variables in the data
C. The color depth of images
D. The speed of computation

Answer: B

Q353. What is the formula for calculating the mean?

A. Sum of values divided by count of values
B. Product of all values
C. Largest value minus smallest value
D. Square root of variance

Answer: A

Q354. What is variance in statistics?

A. The middle value of a dataset
B. A measure of how spread out data points are from the mean
C. The most common value
D. The total count of data points

Answer: B

Q355. What is a function in mathematics?

A. A random process
B. A rule that assigns each input exactly one output
C. A type of matrix
D. A data storage format

Answer: B

Q356. What does the slope of a line represent?

A. The y-intercept
B. The rate of change between two variables
C. The color of the line
D. The length of the line

Answer: B

Q357. What is the difference between a row vector and a column vector?

A. There is no difference
B. A row vector is horizontal while a column vector is vertical
C. Row vectors are larger
D. Column vectors cannot be used in calculations

Answer: B

Q358. What is a conditional probability?

A. The probability of an impossible event
B. The probability of an event occurring given that another event has already occurred
C. A probability greater than one
D. The total probability of all events

Answer: B

Q359. What is an exponent in mathematics?

A. A subtraction operation
B. The number of times a base number is multiplied by itself
C. A type of average
D. A statistical test

Answer: B

Q360. What is a dictionary in Python?

A. An ordered list of elements
B. A collection of key-value pairs
C. A type of loop
D. A mathematical function

Answer: B

Q361. What does the len() function return in Python?

A. The data type of an object
B. The number of items in an object
C. The memory size in bytes
D. The last element of a list

Answer: B

Q362. Which Python library is used for deep learning with dynamic computation graphs?

A. pandas
B. matplotlib
C. PyTorch
D. scikit-learn

Answer: C

Q363. What does the range() function generate in Python?

A. Random numbers
B. A sequence of numbers
C. A list of strings
D. Boolean values

Answer: B

Q364. What is a tuple in Python?

A. A mutable list
B. An immutable ordered sequence of elements
C. A type of dictionary
D. A function definition

Answer: B

Q365. What does the type() function do in Python?

A. Converts data types
B. Returns the data type of an object
C. Creates a new variable
D. Deletes an object

Answer: B

Q366. What is the purpose of the import statement in Python?

A. To export data to files
B. To bring external modules and libraries into the current script
C. To delete variables
D. To create loops

Answer: B

Q367. What does df.describe() do in pandas?

A. Deletes the DataFrame
B. Generates descriptive statistics like mean, std, min, max for numerical columns
C. Sorts the DataFrame
D. Renames columns

Answer: B

Q368. What is a for loop used for in Python?

A. Defining functions
B. Iterating over a sequence of elements
C. Importing libraries
D. Handling exceptions

Answer: B

Q369. What does np.zeros() create in NumPy?

A. An array filled with ones
B. An array filled with zeros
C. An empty array
D. A random array

Answer: B

Q370. What is data transformation in preprocessing?

A. Deleting all data
B. Converting data from one format or structure to another suitable for analysis
C. Printing data on paper
D. Sending data over the internet

Answer: B

Q371. What is the purpose of splitting data into training and testing sets?

A. To reduce dataset size
B. To evaluate model performance on unseen data and detect overfitting
C. To make training faster
D. To remove missing values

Answer: B

Q372. What is a numerical feature in a dataset?

A. A feature containing text labels
B. A feature represented by numbers that can be measured or counted
C. A feature with only two values
D. A feature with missing values

Answer: B

Q373. What does it mean to drop a column in data preprocessing?

A. Moving a column to a different table
B. Removing an entire feature from the dataset
C. Adding new data to a column
D. Sorting a column in descending order

Answer: B

Q374. What is label encoding used for?

A. Adding labels to charts
B. Converting categorical text values into numerical codes
C. Encrypting data for security
D. Splitting data into training sets

Answer: B

Q375. What is the purpose of data validation in preprocessing?

A. Making data look prettier
B. Checking that data meets expected quality standards and constraints before processing
C. Encrypting sensitive data
D. Compressing data files

Answer: B

Q376. What is a feature vector?

A. A type of virus
B. A numerical representation of an object's features used as input to an ML model
C. A graph showing feature importance
D. A physical measurement tool

Answer: B

Q377. What does data integration involve in preprocessing?

A. Splitting data into smaller pieces
B. Combining data from multiple sources into a unified dataset
C. Deleting duplicate records only
D. Changing data types

Answer: B

Q378. Why might you convert text to lowercase during preprocessing?

A. To save storage space
B. To ensure consistent treatment of words regardless of capitalization
C. To make text unreadable
D. To increase processing speed

Answer: B

Q379. What is the purpose of exploratory data analysis?

A. To deploy models to production
B. To understand data patterns, distributions, and relationships before modeling
C. To encrypt sensitive data
D. To compress data files

Answer: B

Q380. What does a histogram display?

A. The relationship between two variables
B. The frequency distribution of a single variable
C. A geographical map
D. A network diagram

Answer: B

Q381. What is the range of a dataset?

A. The average value
B. The difference between the maximum and minimum values
C. The most frequent value
D. The middle value

Answer: B

Q382. What does a negative correlation between two variables mean?

A. Both variables increase together
B. As one variable increases the other tends to decrease
C. The variables are identical
D. There is no relationship

Answer: B

Q383. What is the purpose of a count plot?

A. To show the count of numerical values
B. To show the frequency of each category in a categorical variable
C. To display time series data
D. To plot mathematical functions

Answer: B

Q384. What does df.shape return in pandas?

A. The data types of columns
B. A tuple showing the number of rows and columns in the DataFrame
C. The column names
D. The sum of all values

Answer: B

Q385. What is a missing value indicator in EDA?

A. A value of zero
B. A marker like NaN or null that indicates data was not recorded
C. A negative number
D. The most frequent value

Answer: B

Q386. What is a stacked bar chart used for?

A. Showing 3D data
B. Comparing the composition of categories across different groups
C. Displaying continuous distributions
D. Plotting geographic data

Answer: B

Q387. What is the mean absolute deviation?

A. The square of variance
B. The average of the absolute differences between each value and the mean
C. The maximum deviation from the mean
D. The standard deviation squared

Answer: B

Q388. Why is it important to check the data types of columns during EDA?

A. Data types affect font size in visualizations
B. Incorrect data types can cause errors in computations and analysis, such as numbers stored as strings
C. Data types are irrelevant for analysis
D. All columns should always be strings

Answer: B

Q389. What is the difference between classification and regression?

A. They are the same thing
B. Classification predicts categories while regression predicts continuous numerical values
C. Regression predicts categories while classification predicts numbers
D. Neither involves prediction

Answer: B

Q390. What is a decision boundary in classification?

A. The edge of a dataset
B. The boundary separating different class regions in feature space
C. A programming error
D. The maximum number of features allowed

Answer: B

Q391. What does the term 'prediction' mean in supervised learning?

A. Guessing without any basis
B. The output generated by a trained model for a given input
C. Storing data in a database
D. Visualizing data

Answer: B

Q392. What is a hyperparameter in a machine learning model?

A. A parameter learned from training data
B. A configuration set before training that controls the learning process
C. A type of feature in the dataset
D. The model's final prediction

Answer: B

Q393. What is the goal of minimizing a loss function?

A. To increase errors
B. To find model parameters that produce predictions closest to actual values
C. To maximize data size
D. To remove features

Answer: B

Q394. What is a leaf node in a decision tree?

A. The root of the tree
B. A node at the bottom that contains the final prediction or class label
C. An intermediate splitting node
D. The input layer

Answer: B

Q395. What is gradient descent used for in supervised learning?

A. Sorting data
B. Optimizing model parameters by iteratively moving in the direction of steepest decrease in the loss function
C. Creating visualizations
D. Splitting datasets

Answer: B

Q396. What is multiclass classification?

A. A problem with exactly two output classes
B. A classification task where the model predicts one of three or more possible classes
C. A regression problem
D. An unsupervised learning task

Answer: B

Q397. What is the purpose of the intercept term in linear regression?

A. To multiply features together
B. To represent the predicted value when all features are zero
C. To remove outliers
D. To normalize the data

Answer: B

Q398. What is a training error?

A. An error in the code
B. The difference between the model's predictions and actual values on the training data
C. A hardware malfunction
D. A missing value in the dataset

Answer: B

Q399. What is the wisdom of crowds principle in ensemble learning?

A. Only the best model matters
B. Combining multiple diverse models often produces better predictions than any single model alone
C. Crowds always make wrong decisions
D. Ensembles use only one model

Answer: B

Q400. What does the term 'weak learner' mean in boosting?

A. A model that never makes predictions
B. A model that performs only slightly better than random guessing
C. A model with perfect accuracy
D. A model that uses no data

Answer: B

Q401. What is the primary purpose of averaging predictions in a regression ensemble?

A. To increase variance
B. To reduce prediction errors by smoothing out individual model mistakes
C. To make predictions more extreme
D. To delete outlier predictions

Answer: B

Q402. Which of the following is an example of a bagging algorithm?

A. AdaBoost
B. XGBoost
C. Random Forest
D. Logistic Regression

Answer: C

Q403. What happens to overfitting when you add more trees to a Random Forest?

A. Overfitting always increases dramatically
B. Adding more trees generally does not increase overfitting, it tends to stabilize performance
C. The model becomes simpler
D. All trees become identical

Answer: B

Q404. What is the role of the number of estimators hyperparameter in ensemble methods?

A. It controls the number of features used
B. It specifies how many base models are combined in the ensemble
C. It determines the learning rate
D. It sets the maximum depth of trees

Answer: B

Q405. What is a simple averaging ensemble?

A. A single model that averages features
B. An ensemble that takes the arithmetic mean of predictions from multiple models
C. A feature engineering technique
D. A data preprocessing step

Answer: B

Q406. Why are decision trees commonly used as base learners in ensemble methods?

A. They are the only algorithm available
B. They are fast to train, can model nonlinear relationships, and their high variance is reduced by ensemble aggregation
C. They never overfit
D. They always produce identical results

Answer: B

Q407. What is weighted voting in ensemble classification?

A. All models have equal influence
B. Each model's vote is weighted by its estimated quality or confidence
C. Only the best model votes
D. Weights are applied to features

Answer: B

Q408. What is gradient boosting?

A. An algorithm that uses random sampling like bagging
B. A boosting method that sequentially fits models to the residual errors of previous models
C. A single decision tree algorithm
D. An unsupervised clustering method

Answer: B

Q409. What is the difference between clustering and classification?

A. They are the same task
B. Clustering groups data without predefined labels while classification assigns predefined class labels
C. Classification does not use labels
D. Clustering requires labeled data

Answer: B

Q410. What is the purpose of K-Means initialization?

A. To normalize the data
B. To set the initial positions of cluster centroids before the iterative algorithm begins
C. To determine the number of features
D. To clean the data

Answer: B

Q411. What is market basket analysis?

A. Analyzing shopping cart designs
B. Using association rules to find products frequently purchased together
C. A type of regression
D. A supervised learning technique

Answer: B

Q412. What is the objective of PCA?

A. To increase data dimensions
B. To find directions of maximum variance and reduce dimensionality while preserving information
C. To classify data into groups
D. To remove outliers

Answer: B

Q413. What is the elbow in the elbow method for K-Means?

A. The first cluster
B. The point on the inertia curve where adding more clusters gives diminishing improvement, resembling an elbow shape
C. The last data point
D. The maximum number of features

Answer: B

Q414. What is feature reduction?

A. Adding more features to a dataset
B. Decreasing the number of features while retaining important information
C. Deleting all features
D. Normalizing feature values

Answer: B

Q415. What is a cluster center?

A. The first data point in a cluster
B. The representative point of a cluster, often the mean of all points in the cluster
C. The farthest point from other clusters
D. A data point that belongs to no cluster

Answer: B

Q416. What is the difference between PCA and feature selection?

A. They are identical techniques
B. PCA creates new composite features from original features while feature selection chooses a subset of original features
C. PCA only removes features
D. Feature selection creates new features

Answer: B

Q417. What is an autoencoder's bottleneck layer?

A. The output layer
B. A hidden layer with fewer neurons that forces the network to learn compressed representations of the input
C. The input layer
D. A dropout layer

Answer: B

Q418. What is the purpose of the MiniBatchKMeans algorithm?

A. To increase clustering accuracy
B. To provide a faster approximation of K-Means by using random subsets of data for each iteration
C. To handle text data only
D. To replace DBSCAN

Answer: B

Q419. What is a true positive in classification?

A. The model incorrectly predicts the positive class
B. The model correctly predicts the positive class
C. The model correctly predicts the negative class
D. The model incorrectly predicts the negative class

Answer: B

Q420. What is a true negative in classification?

A. The model correctly predicts the positive class
B. The model incorrectly predicts the positive class
C. The model correctly predicts the negative class
D. The model incorrectly predicts the negative class

Answer: C

Q421. What is a false negative in classification?

A. The model correctly predicts positive
B. The model correctly predicts negative
C. The model incorrectly predicts negative when the actual class is positive
D. The model incorrectly predicts positive when the actual class is negative

Answer: C

Q422. What is the purpose of a holdout test set?

A. To train the model
B. To provide a final unbiased evaluation of the model's performance on unseen data
C. To tune hyperparameters
D. To clean the data

Answer: B

Q423. What does a higher accuracy mean for a classification model?

A. The model makes more errors
B. A larger proportion of all predictions are correct
C. The model is more complex
D. The model trains faster

Answer: B

Q424. What is root mean squared error?

A. The mean of errors
B. The square root of the average of squared differences between predicted and actual values
C. The maximum error
D. The median error

Answer: B

Q425. What does the R-squared value tell us about a regression model?

A. The number of features used
B. The proportion of variance in the target variable explained by the model
C. The training time
D. The number of data points

Answer: B

Q426. Why is it important to evaluate a model on data it has not been trained on?

A. It is not important
B. To check if the model generalizes to new data rather than just memorizing the training data
C. To make training faster
D. To increase the dataset size

Answer: B

Q427. What is a baseline model in model evaluation?

A. The most complex model possible
B. A simple reference model used as a minimum performance benchmark for comparison
C. A model with zero accuracy
D. The final deployed model

Answer: B

Q428. What is the specificity of a classifier?

A. The same as recall
B. The proportion of actual negatives that are correctly identified as negative
C. The total number of predictions
D. The model's complexity level

Answer: B

Q429. What is an interaction feature?

A. A feature from user interactions with a website
B. A new feature created by combining two or more existing features, such as multiplying them together
C. A feature that interacts with the model
D. A type of categorical feature

Answer: B

Q430. What is feature extraction?

A. Removing all features from data
B. Creating new informative features from raw data like extracting day of week from a date
C. A type of model training
D. Deleting columns from a dataset

Answer: B

Q431. Why is domain knowledge valuable in feature engineering?

A. It is not valuable at all
B. Experts can identify meaningful transformations and combinations that algorithms alone might miss
C. Domain knowledge slows down the process
D. It only helps with data collection

Answer: B

Q432. What is a ratio feature?

A. A feature with a fixed ratio
B. A new feature created by dividing one feature by another to capture proportional relationships
C. A feature that counts occurrences
D. A binary feature

Answer: B

Q433. What is the purpose of extracting date components as features?

A. To encrypt date information
B. To create separate features like year, month, day, and day of week that capture temporal patterns
C. To delete date columns
D. To format dates for display

Answer: B

Q434. What is text vectorization?

A. Converting vectors to text
B. Converting text data into numerical feature vectors that ML algorithms can process
C. A text editing operation
D. A type of font transformation

Answer: B

Q435. What is the purpose of indicator (dummy) variables?

A. To indicate errors in data
B. To represent categorical variables as binary columns where each category gets its own column
C. To point to other datasets
D. To indicate missing values only

Answer: B

Q436. What is feature aggregation?

A. Splitting features into smaller parts
B. Summarizing multiple values into a single statistic like count, mean, or sum for grouped data
C. Deleting aggregated data
D. A type of normalization

Answer: B

Q437. What is a count feature?

A. A feature that counts model parameters
B. A feature representing the frequency or count of occurrences of a particular event or category
C. A feature with only integer values
D. The total number of features

Answer: B

Q438. What is the benefit of creating squared features?

A. They reduce all values to zero
B. They allow linear models to capture quadratic (U-shaped or inverted-U) relationships between features and the target
C. They make features easier to read
D. They always improve accuracy

Answer: B

Q439. What is a perceptron in deep learning?

A. A type of dataset
B. The simplest form of neural network with a single neuron that computes a weighted sum of inputs
C. A loss function
D. A type of optimizer

Answer: B

Q440. What is the purpose of the input layer in a neural network?

A. To process data through complex computations
B. To receive and pass the raw input features to the next layer
C. To produce final predictions
D. To apply activation functions

Answer: B

Q441. What is the difference between a deep and a shallow neural network?

A. Deep networks use more data
B. Deep networks have many hidden layers while shallow networks have one or two hidden layers
C. Shallow networks are more accurate
D. There is no difference

Answer: B

Q442. What is the tanh activation function?

A. A linear function
B. A function that outputs values between -1 and 1, shaped like a stretched sigmoid
C. A function that outputs only 0 or 1
D. A function that outputs only positive values

Answer: B

Q443. What is a batch size in neural network training?

A. The total dataset size
B. The number of training samples processed before the model's weights are updated
C. The number of layers in the network
D. The number of output classes

Answer: B

Q444. What is a dense layer in a neural network?

A. A layer with no connections
B. A fully connected layer where every neuron is connected to every neuron in the previous layer
C. A layer that reduces dimensions
D. A layer that only processes images

Answer: B

Q445. What is the purpose of a learning rate schedule?

A. To set a fixed learning rate
B. To adjust the learning rate during training, typically decreasing it to allow finer convergence
C. To increase the learning rate continuously
D. To determine the batch size

Answer: B

Q446. What is the output layer of a neural network?

A. The first layer that receives input
B. The final layer that produces the network's prediction or output
C. Any hidden layer
D. The layer that stores training data

Answer: B

Q447. What is the role of a GPU in deep learning?

A. It stores the training data
B. It accelerates matrix computations needed for training neural networks through parallel processing
C. It provides internet connectivity
D. It manages the operating system

Answer: B

Q448. What is model inference in deep learning?

A. The process of training a model
B. Using a trained model to make predictions on new data
C. Designing the model architecture
D. Cleaning the training data

Answer: B

Q449. What is a stride in CNN convolution?

A. The size of the filter
B. The number of pixels the filter moves across the input for each step
C. The depth of the network
D. The number of filters used

Answer: B

Q450. What is padding in CNNs?

A. Adding noise to images
B. Adding extra pixels around the input borders to control the spatial dimensions of the output
C. Removing pixels from images
D. A type of activation function

Answer: B

Q451. What is average pooling?

A. Computing the average of all network weights
B. Taking the average of values in each pooling window to reduce spatial dimensions
C. Averaging predictions from multiple models
D. Computing the average loss

Answer: B

Q452. What is the vanishing gradient problem specifically in RNNs?

A. Gradients become too large
B. Gradients diminish exponentially over long sequences, making it hard for the network to learn long-range dependencies
C. RNNs do not have gradients
D. Gradients are always stable in RNNs

Answer: B

Q453. What is the generator in a GAN?

A. The part that classifies real from fake
B. The neural network that creates synthetic data samples from random noise
C. A data preprocessing tool
D. The loss function

Answer: B

Q454. What is the purpose of flattening in a CNN?

A. Making images flat
B. Converting multi-dimensional feature maps into a one-dimensional vector to feed into fully connected layers
C. Reducing the number of filters
D. Removing padding from feature maps

Answer: B

Q455. What is global average pooling?

A. Averaging all model parameters
B. Taking the average of each entire feature map to produce a single value per map, replacing fully connected layers
C. A global training strategy
D. Averaging across all training epochs

Answer: B

Q456. What is a feature map in a CNN?

A. A geographical map of features
B. The output produced when a convolutional filter is applied to the input, showing where patterns are detected
C. A list of feature names
D. The model architecture diagram

Answer: B

Q457. What is sequence-to-sequence learning?

A. Learning from one data point to another
B. A framework where the model reads an input sequence and produces an output sequence, used in translation and summarization
C. A type of sorting algorithm
D. Learning sequences in random order

Answer: B

Q458. What is fine-tuning a pre-trained model?

A. Training from scratch on new data
B. Taking a model trained on one task and continuing training on a new related task with a smaller learning rate
C. Removing all layers of a model
D. Only using the model for inference

Answer: B

Q459. What is a vocabulary in NLP?

A. A grammar textbook
B. The set of all unique words or tokens recognized by an NLP model
C. A list of synonyms
D. A type of neural network

Answer: B

Q460. What is text preprocessing in NLP?

A. Writing text for a publication
B. Cleaning and transforming raw text into a format suitable for analysis, including lowercasing, removing punctuation, and tokenizing
C. Compressing text files
D. Printing text documents

Answer: B

Q461. What is a bigram in NLP?

A. A type of neural network
B. A sequence of two consecutive words or characters in text
C. A binary encoding scheme
D. A grammatical rule

Answer: B

Q462. What is the bag-of-words representation?

A. A physical bag containing word cards
B. A text representation that counts word occurrences regardless of order or grammar
C. A neural network for word generation
D. A method for sorting words alphabetically

Answer: B

Q463. What is question answering in NLP?

A. A quiz game format
B. The task of automatically finding or generating answers to questions posed in natural language
C. A type of data collection
D. A text formatting technique

Answer: B

Q464. What is text summarization?

A. Making text longer and more detailed
B. Automatically creating a shorter version of text that retains the key information
C. Translating text to another language
D. Correcting grammar errors

Answer: B

Q465. What is a token in NLP?

A. A coin-like object
B. A basic unit of text such as a word, subword, or character that the model processes
C. A model parameter
D. A type of loss function

Answer: B

Q466. What is spam detection in NLP?

A. Finding physical spam mail
B. Using text classification to automatically identify unwanted or unsolicited messages
C. A network security tool
D. A data cleaning technique

Answer: B

Q467. What is text generation in NLP?

A. Manually typing text
B. Automatically producing coherent and contextually appropriate text using language models
C. Copying text from one file to another
D. Formatting text for printing

Answer: B

Q468. What is image preprocessing in computer vision?

A. Taking photographs
B. Preparing images for analysis through operations like resizing, normalization, and color conversion
C. Printing images
D. Compressing image file sizes only

Answer: B

Q469. What is a bounding box in object detection?

A. A box used for shipping
B. A rectangle drawn around a detected object specifying its location in the image
C. A type of convolutional filter
D. A border around the entire image

Answer: B

Q470. What is the difference between a color image and a grayscale image?

A. They have the same number of channels
B. Color images have three channels (RGB) while grayscale images have a single channel representing brightness
C. Grayscale images are larger in file size
D. Color images have only one channel

Answer: B

Q471. What is image annotation in computer vision?

A. Writing captions for social media
B. Labeling images with class labels, bounding boxes, or pixel masks to create training data for CV models
C. Editing image colors
D. Compressing images

Answer: B

Q472. What is transfer learning commonly used for in computer vision?

A. Transferring images between devices
B. Using a model pre-trained on a large dataset and adapting it to a new visual task with less data
C. Transferring data between databases
D. Moving files between folders

Answer: B

Q473. What does image normalization do?

A. Makes all images the same size
B. Scales pixel values to a standard range like 0 to 1 or standardizes them using mean and standard deviation
C. Removes noise from images
D. Converts images to text

Answer: B

Q474. What is edge detection in image processing?

A. Detecting the physical edges of a monitor
B. Identifying boundaries between different regions in an image where pixel intensity changes sharply
C. Finding the borders of the image file
D. A type of data augmentation

Answer: B

Q475. What is image cropping in computer vision?

A. Growing images like crops
B. Selecting and extracting a rectangular region of interest from an image
C. Deleting all images
D. Rotating images

Answer: B

Q476. What is image flipping as a data augmentation technique?

A. Inverting pixel values
B. Mirroring an image horizontally or vertically to create new training samples
C. Converting the image format
D. Changing image resolution

Answer: B

Q477. What is a pre-trained model like ResNet or VGG used for?

A. Only for the task it was originally trained on
B. As a feature extractor or starting point for transfer learning on new visual tasks
C. It cannot be reused
D. Only for generating images

Answer: B

Q478. What is a cluster in distributed computing?

A. A group of similar data points
B. A collection of interconnected computers working together to process data
C. A type of database index
D. A sorting algorithm

Answer: B

Q479. What is the Map phase in MapReduce?

A. Creating geographic maps
B. Processing input data and emitting key-value pairs for each record
C. Reducing data to a single value
D. Storing data in memory

Answer: B

Q480. What is the Reduce phase in MapReduce?

A. Deleting unnecessary data
B. Aggregating all intermediate values associated with the same key to produce final results
C. Creating new data points
D. Mapping data to categories

Answer: B

Q481. What is a NoSQL database?

A. A database that does not store data
B. A non-relational database designed for flexible schemas, horizontal scaling, and specific data models like document, key-value, or graph
C. A SQL database with extra features
D. A database only for numerical data

Answer: B

Q482. What is real-time data processing?

A. Processing data once a year
B. Processing data immediately or within milliseconds as it arrives, enabling instant insights and actions
C. Processing data in monthly batches
D. Storing data for future processing

Answer: B

Q483. What is data replication in distributed systems?

A. Deleting duplicate data
B. Storing copies of data on multiple nodes to ensure availability and fault tolerance
C. Compressing data to save space
D. Encrypting data for security

Answer: B

Q484. What is a data pipeline?

A. A physical pipe for data cables
B. A series of automated steps that move and transform data from source systems to destination systems
C. A single database query
D. A type of network cable

Answer: B

Q485. What is horizontal scaling?

A. Making a single server larger
B. Adding more machines to a system to distribute the workload across multiple nodes
C. Increasing CPU speed
D. Adding more RAM to one server

Answer: B

Q486. What is a Spark RDD?

A. A type of hard drive
B. A Resilient Distributed Dataset, an immutable distributed collection of objects that can be processed in parallel
C. A relational database design
D. A random data distribution

Answer: B

Q487. What is cloud computing in the context of big data?

A. Computing in cloudy weather
B. Using remote servers hosted on the internet to store, manage, and process data instead of local infrastructure
C. A type of encryption
D. A programming language

Answer: B

Q488. What is the goal of MLOps?

A. To replace data scientists
B. To streamline the process of deploying, monitoring, and maintaining ML models in production
C. To eliminate the need for testing
D. To make models train faster only

Answer: B

Q489. What is a model endpoint in deployment?

A. The end of model training
B. A URL or API where a deployed model receives input data and returns predictions
C. The last layer of a neural network
D. The final evaluation metric

Answer: B

Q490. What is containerization in MLOps?

A. Putting computers in containers
B. Packaging a model with its dependencies and environment into a container for consistent deployment across different platforms
C. A type of data storage
D. A compression technique

Answer: B

Q491. What is the purpose of logging in ML systems?

A. Cutting trees for lumber
B. Recording system events, predictions, inputs, and errors to enable debugging, monitoring, and auditing of ML systems
C. Slowing down the system
D. Only tracking training loss

Answer: B

Q492. What is batch inference?

A. Training in batches
B. Running predictions on a large collection of data at once, typically on a schedule, rather than one at a time
C. Real-time prediction for each request
D. A type of model training

Answer: B

Q493. What is the purpose of health checks for deployed ML models?

A. Checking the health of data scientists
B. Automated checks that verify the model service is running, responsive, and functioning correctly
C. Only checking hardware temperature
D. Measuring model accuracy once

Answer: B

Q494. What is a rollback in model deployment?

A. Rolling the model forward
B. Reverting to a previous model version when the new version performs poorly or causes issues
C. Restarting the entire system
D. Deleting all model versions

Answer: B

Q495. What is an ML experiment?

A. A chemistry lab experiment
B. A specific model training run with defined hyperparameters, data, and code that produces measurable results
C. A random guess
D. A data collection process

Answer: B

Q496. What is latency in the context of model serving?

A. The accuracy of the model
B. The time it takes from receiving an input to returning a prediction
C. The size of the model
D. The training time

Answer: B

Q497. What is the purpose of automated testing in MLOps?

A. To test hardware components
B. To automatically verify that ML code, data, and models meet quality standards before deployment
C. Testing is not important for ML
D. To slow down the deployment process

Answer: B

Q498. What is responsible AI?

A. AI that responds quickly
B. The practice of developing and deploying AI systems that are fair, transparent, accountable, and safe
C. AI that costs less money
D. AI that only works for governments

Answer: B

Q499. What is an example of AI bias in hiring?

A. AI that works faster than humans
B. An AI system that unfairly favors candidates of a particular gender or race based on biased training data
C. AI that processes more applications
D. AI that costs less than human recruiters

Answer: B

Q500. Why is human oversight important in AI decision-making?

A. Humans are always faster than AI
B. To ensure AI decisions are reviewed, correctable, and accountable, especially in high-stakes situations
C. Human oversight is not necessary
D. AI should make all decisions independently

Answer: B

Q501. What is data consent in AI ethics?

A. Agreeing to pay for data
B. Obtaining permission from individuals before collecting, using, or sharing their personal data for AI systems
C. Data consent is not required
D. Only companies need to give consent

Answer: B

Q502. What is the digital divide in the context of AI?

A. A type of data splitting technique
B. The gap between those who have access to AI technology and those who do not, often along socioeconomic or geographic lines
C. A coding technique
D. A security vulnerability

Answer: B

Q503. What does the term 'black box' mean in AI ethics?

A. A physical black box device
B. An AI system whose internal decision-making process is not understandable or explainable to users
C. A box for storing AI hardware
D. A type of test environment

Answer: B

Q504. What is the potential impact of AI on employment?

A. AI has no effect on jobs
B. AI can automate certain tasks, potentially displacing some jobs while creating new roles that require human-AI collaboration skills
C. AI will eliminate all human jobs
D. AI only creates new jobs

Answer: B

Q505. What is the purpose of AI ethics guidelines?

A. To slow down AI development
B. To provide principles and frameworks for developing AI systems that are beneficial, fair, and minimize harm to society
C. To make AI more expensive
D. To prevent all AI research

Answer: B

Q506. What is surveillance using AI technology?

A. AI monitoring computer performance
B. Using AI systems to monitor, track, or analyze people's activities, raising privacy and civil liberties concerns
C. A type of AI training method
D. A data backup process

Answer: B

Q507. Why is diversity important in AI development teams?

A. It has no impact on AI quality
B. Diverse teams are more likely to identify potential biases, consider varied perspectives, and build AI systems that work fairly for all populations
C. Only technical skills matter
D. Diversity slows down development

Answer: B

Q508. What is the absolute value of a number?

A. The number multiplied by itself
B. The non-negative distance of a number from zero on the number line
C. The reciprocal of the number
D. The square root of the number

Answer: B

Q509. What is the purpose of data type conversion in preprocessing?

A. To delete data
B. To change values from one data type to another so they can be processed correctly by algorithms
C. To increase file size
D. To add new columns

Answer: B

Q510. What is a chatbot in NLP?

A. A physical robot that chats
B. A software application that simulates human conversation using natural language processing
C. A chat room for programmers
D. A tool for compressing text files

Answer: B

Medium Questions

510 questions

Q511. The Turing Test is used to evaluate:

A. Whether a machine can exhibit intelligent behavior
B. The total memory capacity of a given system
C. How quickly a processor handles computations
D. How much bandwidth a network can support

Answer: A

Q512. Which of the following best describes Deep Learning?

A. A relational system for storing structured data
B. A paradigm for writing event-driven programs
C. A subset of ML using multi-layer neural networks
D. A protocol for routing data between networks

Answer: C

Q513. In reinforcement learning, an agent learns by:

A. Memorizing all possible training inputs
B. Receiving rewards or penalties for actions
C. Reading through a set of labeled data points
D. Clustering together groups of similar data

Answer: B

Q514. Which of the following is a descriptive analytics technique?

A. Predicting future trends
B. Real-time optimization
C. Prescribing actions
D. Summarizing historical data

Answer: D

Q515. What is the difference between AI and ML?

A. AI is a subset of ML
B. They are the same thing
C. They are unrelated fields
D. ML is a subset of AI

Answer: D

Q516. Which type of analytics answers 'What will happen?'

A. Prescriptive Analytics
B. Descriptive Analytics
C. Predictive Analytics
D. Diagnostic Analytics

Answer: C

Q517. An expert system in AI uses:

A. A knowledge base and inference engine
B. Only probabilistic fuzzy logic rules
C. Only evolutionary genetic algorithms
D. Only deep artificial neural networks

Answer: A

Q518. Which of these is a weak AI system?

A. A chess-playing program
B. A general-purpose intellect
C. A conscious thinking machine
D. A fully self-aware robot

Answer: A

Q519. What is the role of a training set in ML?

A. To deploy the model into production
B. To visualize the model's final results
C. To test the final trained model output
D. To train the model to learn patterns

Answer: D

Q520. Prescriptive analytics is used to:

A. Describe past events only
B. Store data more efficiently
C. Recommend actions to take
D. Detect anomalies in data

Answer: C

Q521. The gradient of a function points in the direction of:

A. Steepest descent
B. Random direction
C. Steepest ascent
D. Zero change

Answer: C

Q522. Bayes' theorem relates:

A. Prior and posterior probabilities
B. Only matrix decompositions
C. Only the mean and variance
D. Only derivative computations

Answer: A

Q523. The eigenvalue equation is Av = ;·`. What is ;³ð

A. Eigenvector
B. Trace
C. Determinant
D. Eigenvalue

Answer: D

Q524. What is the chain rule used for in calculus?

A. Computing joint probabilities
B. Adding matrices element-wise
C. Differentiating composite functions
D. Sorting data sequentially

Answer: C

Q525. Standard deviation measures:

A. Spread of data around the mean
B. The central tendency of values
C. The minimum observed value
D. The maximum observed value

Answer: A

Q526. A positive definite matrix has:

A. All negative eigenvalues
B. Mixed sign eigenvalues
C. All zero eigenvalues
D. All positive eigenvalues

Answer: D

Q527. In probability, two events are independent if:

A. P(A|B) = P(B)
B. P(A")B) = P(A) * P(B)
C. P(A")B) = P(A) + P(B)
D. P(A")B) = 0

Answer: B

Q528. The determinant of a 2x2 matrix [[a,b],[c,d]] is:

A. ab + cd
B. ab - cd
C. ad - bc
D. ad + bc

Answer: C

Q529. A convex function has the property that:

A. It is always an increasing function
B. It has no minimum point at all
C. Any local minimum is also a global minimum
D. It has many distinct local minima

Answer: C

Q530. The covariance between two identical variables equals:

A. A value of exactly zero
B. The variance of that variable
C. A value of negative one
D. A value of exactly one

Answer: B

Q531. What is broadcasting in NumPy?

A. A networking feature for sending data to multiple receivers
B. Automatic expansion of arrays with different shapes for arithmetic
C. A type of sorting algorithm for ordering array elements
D. A logging mechanism for recording runtime error messages

Answer: B

Q532. Which scikit-learn class is used for train-test splitting?

A. SplitData
B. DataSplitter
C. train_test_split
D. TestTrainDivide

Answer: C

Q533. What does df.groupby() do in pandas?

A. Renames columns using a provided mapping dict
B. Sorts the DataFrame by index values in order
C. Filters rows based on a boolean condition mask
D. Groups data by one or more columns for aggregation

Answer: D

Q534. Which pandas method fills missing values?

A. fillna()
B. isna()
C. notna()
D. dropna()

Answer: A

Q535. What is a Python generator?

A. A built-in immutable data type like a list
B. A function that yields values lazily using yield
C. A special class constructor for initialization
D. A comparison-based sorting algorithm method

Answer: B

Q536. How do you perform matrix multiplication in NumPy?

A. np.multiply(A, B) function
B. A + B element-wise addition
C. A * B element-wise product
D. np.dot(A, B) or A @ B operator

Answer: A

Q537. What does the pandas method .apply() do?

A. Deletes specified rows from the DataFrame index
B. Merges two DataFrames on a common column key
C. Adds a new column to the end of a DataFrame
D. Applies a function along an axis of a DataFrame

Answer: D

Q538. Which visualization library is built on top of Matplotlib?

A. Pygal
B. Bokeh
C. Seaborn
D. Plotly

Answer: C

Q539. What is the purpose of np.reshape()?

A. To sort the elements of an array in ascending order
B. To change the shape of an array without changing its data
C. To filter elements of an array by a condition mask
D. To delete selected elements from an existing array

Answer: B

Q540. What does pickle do in Python?

A. Manages relational database connections
B. Performs numerical matrix calculations
C. Creates interactive charts and plots
D. Serializes and deserializes Python objects

Answer: D

Q541. What is one-hot encoding?

A. Encrypting data columns for secure transmission
B. Compressing data files for reduced disk usage
C. Converting categorical variables into binary vectors
D. Normalizing continuous numerical data columns

Answer: C

Q542. What is the difference between normalization and standardization?

A. Normalization scales to [0,1], standardization scales to zero mean and unit variance
B. Normalization is always the superior approach for every ML algorithm
C. They are identical processes producing the exact same transformed outputs
D. Standardization is only applicable to integer-valued feature columns

Answer: A

Q543. What is label encoding?

A. Assigning numerical values to categorical labels
B. Encrypting labels for secure data storage
C. Adding new labels to the feature space
D. Removing labels from a dataset entirely

Answer: A

Q544. What is imputation?

A. Adding new columns to expand the feature set
B. Replacing missing values with estimated values
C. Deleting every row containing missing values
D. Sorting data by ascending numerical order

Answer: B

Q545. Why is feature scaling important for ML algorithms?

A. It automatically removes all outlier data point values
B. It adds new synthetic features to expand the model
C. It makes the overall dataset physically larger in storage
D. It ensures all features contribute equally regardless of their scale

Answer: D

Q546. What is data augmentation?

A. Encrypting data columns for privacy and security
B. Deleting redundant records from the training set
C. Compressing data files to reduce storage overhead
D. Creating new training data by modifying existing data

Answer: D

Q547. What is the purpose of the train-test split?

A. To evaluate model performance on unseen data
B. To increase the total size of the full dataset
C. To clean and preprocess the raw input data
D. To visualize the distribution of data points

Answer: A

Q548. What is an outlier?

A. A value that is completely missing or absent
B. A feature that stores categorical text labels
C. A data point significantly different from others
D. A record that has been duplicated in error

Answer: C

Q549. What does the z-score standardization formula compute?

A. (x - min) / (max - min) range
B. x divided by the max value
C. (x - mean) / standard deviation
D. log of the value x

Answer: C

Q550. What is web scraping in data collection?

A. Designing interactive web page layouts
B. Testing web application load performance
C. Building responsive websites from scratch
D. Extracting data from websites programmatically

Answer: D

Q551. What is a heatmap used for in EDA?

A. Showing temperature measurements from a sensor only
B. Visualizing the magnitude of values in a matrix using colors
C. Building multi-layer neural network architectures
D. Creating three-dimensional surface projection plots

Answer: B

Q552. Skewness in a distribution refers to:

A. Asymmetry of the distribution around the mean
B. The total number of distribution peaks
C. The height of the data distribution
D. The overall width of the distribution

Answer: C

Q553. What is a pair plot (pairplot) in Seaborn?

A. A single scatter plot between exactly two variables
B. A grid of plots showing pairwise relationships between variables
C. A grouped bar chart comparing category counts
D. A circular pie chart showing value proportions

Answer: B

Q554. Kurtosis measures:

A. The arithmetic mean
B. The tailedness of a distribution
C. The center of the distribution
D. The measure of skewness

Answer: C

Q555. What is the Interquartile Range (IQR)?

A. Mean minus Median
B. Max - Min range
C. Standard deviation * 2
D. Q3 - Q1 difference

Answer: C

Q556. A violin plot combines which two visualizations?

A. Box plot and KDE plot
B. Pie chart and histogram
C. Scatter plot and bar chart
D. Line chart and area chart

Answer: A

Q557. What does the value_counts() method do in pandas?

A. Returns the frequency of unique values in a Series
B. Sums together all values stored in a Series
C. Counts the total number of rows in a Series
D. Counts the number of null values in a Series

Answer: B

Q558. A QQ plot is used to:

A. Build linear regression prediction models
B. Plot quarterly financial data over time
C. Create grouped and stacked bar charts
D. Check if data follows a particular distribution

Answer: D

Q559. What is a kernel density estimation (KDE) plot?

A. A circular chart showing proportions
B. A variant of the standard scatter plot
C. A smoothed continuous version of a histogram
D. A type of stacked or grouped bar chart

Answer: C

Q560. What is the purpose of a log transformation in EDA?

A. To reduce skewness and make data more normally distributed
B. To change the data types of columns in the table
C. To encrypt data columns for compliance and security
D. To completely delete all outlier values from data

Answer: A

Q561. What is the cost function in linear regression?

A. Hinge loss function
B. Log loss function
C. Cross-entropy loss
D. Mean Squared Error (MSE)

Answer: B

Q562. The sigmoid function maps values to:

A. Range between 0 and 1
B. Only integers
C. Range between -1 and 1
D. Any real number

Answer: A

Q563. What is regularization in supervised learning?

A. Increasing overall model complexity further
B. Adding more training data to the model
C. Adding a penalty term to prevent overfitting
D. Removing unnecessary features from the model

Answer: C

Q564. What is the difference between L1 and L2 regularization?

A. L1 uses absolute values and can produce sparse models; L2 uses squared values
B. L1 and L2 are completely identical techniques
C. L2 regularization always produces sparse model weights
D. L1 regularization is always the better choice overall

Answer: D

Q565. What is a Support Vector Machine (SVM)?

A. An algorithm that finds the optimal hyperplane to separate classes
B. A linear dimensionality reduction projection method
C. A type of deep feedforward artificial neural network
D. An unsupervised density-based clustering algorithm

Answer: A

Q566. What is the bias-variance tradeoff?

A. Balancing model simplicity (bias) against sensitivity to training data (variance)
B. Choosing between running computation on a CPU versus a GPU device
C. Selecting the right proportions for training and test data splits
D. Choosing an appropriate learning rate for gradient optimization

Answer: A

Q567. In decision trees, what is information gain?

A. The reduction in entropy after splitting on a feature
B. The overall accuracy of the trained model
C. The total number of leaf nodes in the tree
D. The current depth of the decision tree

Answer: B

Q568. What is cross-validation?

A. Evaluating using only the training data subset
B. Splitting data into multiple folds to train and validate the model
C. Training the model on all available data at once
D. Evaluating using only the test data subset

Answer: D

Q569. What is the purpose of the softmax function?

A. Converting raw scores into probabilities for multi-class classification
B. Performing binary classification tasks on labeled data only
C. Performing unsupervised clustering of unlabeled data points
D. Performing regression to predict continuous numeric values

Answer: A

Q570. What is Naive Bayes based on?

A. A deep learning based neural network model
B. Bayes' theorem with an assumption of feature independence
C. An iterative gradient descent optimization method
D. An ensemble of many decision tree classifiers

Answer: B

Q571. What is the difference between bagging and boosting?

A. Bagging and boosting are completely identical approaches
B. Boosting is always faster than bagging in every scenario
C. Bagging always performs better than boosting overall
D. Bagging trains models independently in parallel; boosting trains them sequentially on errors

Answer: C

Q572. In AdaBoost, how are misclassified samples handled?

A. They are completely removed from the data
B. Their weights are increased for the next iteration
C. Their weights are decreased in the next round
D. They are duplicated in the training dataset

Answer: C

Q573. What is stacking in ensemble learning?

A. Using only a single standalone model for all predictions
B. Random feature selection from the available input columns
C. Using a meta-model to combine predictions from multiple base models
D. Data augmentation to increase the training set size

Answer: C

Q574. What is the out-of-bag (OOB) error in Random Forest?

A. The overall training error computed on the full training dataset
B. The test error computed on a held-out independent test set
C. The validation error from a separate cross-validation fold
D. Error estimated using samples not included in each tree's bootstrap sample

Answer: D

Q575. Gradient Boosting minimizes the loss function by:

A. Removing the least informative features iteratively
B. Increasing the overall size of the training dataset
C. Adding trees that fit the negative gradient of the loss
D. Random bootstrap sampling of the original dataset

Answer: C

Q576. What is the max_features parameter in Random Forest?

A. The minimum number of samples per leaf node
B. The number of features to consider at each split
C. The maximum depth of each individual tree
D. The maximum total number of trees in the forest

Answer: C

Q577. Why does ensemble learning generally outperform single models?

A. It reduces variance and/or bias by combining diverse models
B. It always uses significantly less training data overall
C. It is always computationally faster than a single model
D. It always requires fewer input features to train on

Answer: A

Q578. What is a weak learner?

A. A model that performs slightly better than random guessing
B. A model that consistently achieves 100% accuracy
C. A model that never makes any classification errors
D. A model with absolutely no trainable parameters

Answer: A

Q579. In XGBoost, what is the purpose of the learning rate?

A. It controls the contribution of each tree to shrink step size
B. It directly determines the maximum depth of each tree
C. It selects which features are used at each tree split
D. It explicitly sets the total number of trees in the forest

Answer: A

Q580. What is voting in ensemble methods?

A. Training a single standalone model on the complete dataset
B. Engineering new derived features from the raw input data
C. Combining predictions by having each model vote on the outcome
D. Removing outlier data points from the training samples

Answer: C

Q581. How does the elbow method work for K-Means?

A. Select the value of K completely at random from a uniform range
B. Use the largest possible K equal to the number of data points
C. Plot inertia vs K and find where adding clusters gives diminishing returns
D. Always use K equal to three regardless of the dataset size

Answer: C

Q582. What is the silhouette score?

A. The Euclidean distance between every pair of cluster centroids
B. A measure of how similar a point is to its own cluster vs neighboring clusters
C. The total number of clusters found by the clustering algorithm
D. The total number of iterations before the algorithm converges

Answer: B

Q583. DBSCAN stands for:

A. Data-Based Statistical Clustering Algorithm
B. Database Scan Algorithm Notation
C. Density-Based Spatial Clustering of Applications with Noise
D. Deep Binary Sorting Classification Algorithm

Answer: D

Q584. What advantage does DBSCAN have over K-Means?

A. It only works with fully labeled and supervised training data
B. It can find arbitrarily shaped clusters and doesn't require specifying K
C. It always gives strictly better results than all alternatives
D. It is always computationally faster than every other algorithm

Answer: B

Q585. In PCA, principal components are:

A. Orthogonal directions of maximum variance
B. Cluster center points
C. Randomly chosen directions
D. Features that contain missing values

Answer: B

Q586. What is hierarchical clustering?

A. A supervised regression technique for numerical prediction
B. A deep neural network architecture for image recognition
C. Building a tree of clusters by merging or splitting iteratively
D. Using K-Means clustering multiple times in succession

Answer: C

Q587. What is the Apriori algorithm used for?

A. Frequent itemset mining and association rules
B. Dimensionality reduction projections
C. Numerical regression for predictions
D. Supervised classification with labeled data

Answer: A

Q588. What is t-SNE used for?

A. Visualizing high-dimensional data in 2D or 3D
B. Imputing missing data point values
C. Training deep neural network models
D. Selecting the most relevant features

Answer: A

Q589. In association rules, what does 'support' measure?

A. The conditional confidence level of the rule
B. The overall predictive accuracy of the rule
C. The proportion of transactions containing an itemset
D. The total number of features in the dataset

Answer: C

Q590. What is the difference between agglomerative and divisive hierarchical clustering?

A. Neither method uses a hierarchy
B. Agglomerative is bottom-up; divisive is top-down
C. Agglomerative is top-down; divisive is bottom-up
D. The two methods are completely identical

Answer: C

Q591. What is the F1-score?

A. The arithmetic mean of precision and recall
B. The product of precision and recall values
C. The harmonic mean of precision and recall
D. The square of the model accuracy

Answer: A

Q592. What does the ROC curve plot?

A. Training error values plotted against validation error values
B. True Positive Rate vs False Positive Rate at various thresholds
C. Precision values plotted against corresponding recall values
D. Training accuracy values plotted against the loss values

Answer: B

Q593. What does AUC-ROC measure?

A. Only the precision metric of the trained classification model
B. Only the accuracy metric of the trained classification model
C. Only the recall metric of the trained classification model
D. The overall ability of the model to discriminate between classes

Answer: A

Q594. What is K-fold cross-validation?

A. Dividing data into K folds and using each fold as validation once
B. Training the model completely without any validation step
C. Using all available data solely for the training process
D. Using only two simple train-test splits for evaluation

Answer: B

Q595. When is recall more important than precision?

A. When simple accuracy alone is a sufficient metric
B. When the dataset classes are perfectly balanced overall
C. When false positives are very costly to the outcome
D. When missing positive cases is costly, like disease detection

Answer: C

Q596. What is stratified cross-validation?

A. Performing completely random splitting of the data
B. Ignoring overall class balance during data splitting
C. Using only the majority class for model training purposes
D. K-fold CV that preserves the class distribution in each fold

Answer: B

Q597. What is the log loss (cross-entropy loss)?

A. A feature selection method based on mutual information scores
B. The exact same metric as mean squared error for regression
C. A metric specifically designed for evaluating cluster quality
D. A loss function that penalizes confident wrong predictions heavily

Answer: D

Q598. What is the purpose of a learning curve?

A. To select the most important features from the training data
B. To diagnose overfitting or underfitting by plotting performance vs training size
C. To clean and preprocess the raw input data before training
D. To learn about new machine learning algorithms from documentation

Answer: B

Q599. What is the difference between micro and macro averaging?

A. Micro averaging is always a better choice than macro averaging
B. Micro aggregates all instances globally; macro averages per-class metrics equally
C. Micro and macro averaging are completely identical approaches
D. Macro averaging ignores the overall class distribution entirely

Answer: A

Q600. What is Mean Absolute Error (MAE)?

A. The single maximum prediction error value
B. The median of all prediction errors
C. The average of absolute differences between predicted and actual values
D. The average of all squared differences between predictions

Answer: A

Q601. What is the purpose of polynomial features?

A. To remove unnecessary features
B. To normalize the data values
C. To handle all missing values
D. To capture non-linear relationships by creating powers and interactions of features

Answer: A

Q602. What is the Variance Inflation Factor (VIF)?

A. An indicator for missing feature values
B. A measure of multicollinearity between features
C. A metric for overall model accuracy
D. A particular feature scaling method

Answer: D

Q603. What is target encoding?

A. Encoding categories using one-hot vector representation
B. Replacing categories with the mean of the target variable for that category
C. Removing all categorical features from dataset
D. Encoding using binary representation of categories

Answer: A

Q604. Why is feature scaling important for KNN?

A. KNN only works well with categorical type features
B. KNN uses distance calculations that are affected by feature magnitudes
C. Feature scaling actually slows down KNN
D. KNN actually ignores all feature scales

Answer: A

Q605. What is a lag feature in time series?

A. A feature with all of its values missing
B. A feature created from previous time steps' values
C. A feature that holds only binary zero-one
D. A feature that is computationally slow to run

Answer: B

Q606. What is the purpose of log transformation on features?

A. To deliberately add more noise to features
B. To reduce skewness and handle multiplicative relationships
C. To intentionally increase the skewness of data
D. To remove all features entirely from the dataset

Answer: B

Q607. What is feature hashing?

A. Encrypting feature values for secure storage and privacy compliance
B. Mapping high-dimensional features to a fixed-size vector using a hash function
C. Sorting features in the dataset by their statistical importance
D. Removing duplicate feature entries from the data table columns

Answer: B

Q608. What are rolling window features?

A. Features that are created without any time component at all
B. Simple binary indicator features only
C. Static features with no temporal aspect
D. Features computed from a sliding window over time series data like rolling mean

Answer: C

Q609. What is the purpose of the chi-squared test in feature selection?

A. To test the independence between categorical features and the target
B. To normalize all features to a common range
C. To impute missing values in the features
D. To create entirely new features from scratch

Answer: B

Q610. What is domain knowledge in feature engineering?

A. Using only fully automated feature methods
B. Deliberately ignoring the problem context entirely
C. Creating features completely at random each time
D. Using expertise about the problem area to create meaningful features

Answer: C

Q611. What is the vanishing gradient problem?

A. Gradients become very small in early layers, making training slow
B. Gradients become extremely large causing numerical overflow
C. The model has far too many hidden layers to train
D. The training process converges far too quickly to minimum

Answer: A

Q612. What is dropout?

A. Deleting training data samples to reduce dataset size
B. Randomly deactivating neurons during training to prevent overfitting
C. Reducing the learning rate schedule over the epochs
D. Removing network layers permanently from the architecture

Answer: B

Q613. What is batch normalization?

A. A loss function for measuring classification prediction error
B. A preprocessing step applied only to the raw input data
C. Normalizing the inputs of each layer to stabilize and accelerate training
D. A type of non-linear activation function for hidden layers

Answer: C

Q614. What is the difference between a batch and a mini-batch?

A. Batch and mini-batch gradient descent are completely identical
B. A batch is the full dataset; a mini-batch is a subset used per iteration
C. A mini-batch is always larger than a full batch of data
D. A batch always contains exactly one single data sample only

Answer: A

Q615. What is transfer learning?

A. Moving files between different computer file systems
B. Transferring data records between separate databases
C. Using a pre-trained model on a new but related task
D. Training every model from scratch on random weights

Answer: C

Q616. What is the purpose of the learning rate?

A. It determines the total number of hidden layers in the model
B. It explicitly sets the mini-batch size during each epoch
C. It defines the specific neural network architecture layout
D. It controls the step size of weight updates during optimization

Answer: D

Q617. What is the exploding gradient problem?

A. The overall training process runs far too slowly
B. Gradients become extremely small, causing no weight changes
C. The model severely underfits the training data distribution
D. Gradients become extremely large, causing unstable weight updates

Answer: D

Q618. What is weight initialization and why is it important?

A. Setting initial weights to enable effective training and avoid gradient problems
B. Weights are assigned their final values only after training
C. Weights are always initialized to exactly zero for every neuron
D. The initialization strategy has no effect on the training process

Answer: A

Q619. What is the Adam optimizer?

A. A non-linear activation function like sigmoid or ReLU
B. An adaptive learning rate optimizer combining momentum and RMSProp
C. A cross-entropy loss function for classification training
D. A type of multi-layer feedforward neural network architecture

Answer: B

Q620. What is a softmax output layer used for?

A. Feature extraction from the input data
B. Binary classification with only two output classes
C. Regression for continuous value prediction
D. Multi-class classification producing a probability distribution

Answer: D

Q621. What is the Transformer architecture?

A. A model based on self-attention mechanisms without recurrence
B. A generative adversarial network for data synthesis
C. A density-based clustering algorithm for grouping
D. A standard convolutional neural network for image tasks

Answer: A

Q622. What is self-attention?

A. A cross-entropy loss function that measures classification prediction error
B. A type of spatial pooling operation that reduces feature map dimensions
C. A mechanism relating different positions within a sequence to compute representations
D. A data augmentation method that generates synthetic training examples

Answer: C

Q623. What is a Variational Autoencoder (VAE)?

A. A linear regression model for numeric outputs
B. A generative model that learns a probabilistic latent space
C. A standard deterministic autoencoder without sampling
D. A supervised classifier for categorical prediction

Answer: B

Q624. What is the difference between GRU and LSTM?

A. GRU is always more accurate than LSTM on all tasks
B. GRU and LSTM are completely identical architectures
C. GRU has fewer gates (2 vs 3) and is computationally simpler
D. LSTM has fewer learnable parameters than the GRU

Answer: D

Q625. What is 1x1 convolution used for?

A. Increasing the spatial dimensions of the feature maps in the network
B. Changing the number of channels and adding non-linearity without changing spatial dimensions
C. Performing spatial pooling to reduce the feature map dimensions
D. Applying batch normalization across all the feature map channels

Answer: A

Q626. What is depthwise separable convolution?

A. Performing a pooling operation to reduce the spatial feature dimensions
B. Splitting convolution into depthwise and pointwise operations to reduce computation
C. Applying a normalization technique to stabilize the training process
D. Performing a standard full convolution operation on the input feature maps

Answer: A

Q627. What is the encoder-decoder architecture?

A. A cross-entropy loss function for measuring prediction quality
B. A type of generative adversarial network with two competing agents
C. A structured data format for storing model configuration files
D. A structure where an encoder compresses input and a decoder generates output

Answer: D

Q628. What is data parallelism in deep learning?

A. A type of batch normalization applied across distributed nodes
B. Distributing training data across multiple GPUs, each running a model copy
C. Reducing the overall size of the training data by sampling
D. Using a single GPU for all training and inference computation

Answer: B

Q629. What is a residual network (ResNet)?

A. A very shallow network with only a single hidden layer
B. A standard recurrent neural network for sequences
C. A deep network using skip connections that add input to output of layers
D. A network architecture with no inter-layer connections

Answer: C

Q630. What is the purpose of attention mechanisms?

A. To allow models to focus on relevant parts of the input
B. To increase the mini-batch size during training
C. To initialize the network weights before training
D. To reduce the learning rate over training epochs

Answer: A

Q631. What is TF-IDF?

A. A recurrent neural network variant designed for sequential text modeling
B. A structured query language for relational database data retrieval
C. A multi-layer deep feedforward neural network architecture for classification
D. A weighting scheme reflecting word importance in a document relative to a corpus

Answer: D

Q632. What is Word2Vec?

A. A simple word frequency counter for bag-of-words features
B. A technique learning dense vector representations of words from context
C. A language dictionary for looking up word definitions
D. A rule-based grammar checker for text correction

Answer: B

Q633. What is the bag-of-words model?

A. Representing text with full word order preserved
B. A syntactic parsing technique for sentences
C. Representing text as the frequency of words, ignoring order
D. A neural language model using word embeddings

Answer: C

Q634. What is BERT?

A. A rule-based expert system for language processing
B. A generative adversarial network designed for text data
C. A pre-trained transformer for bidirectional language understanding
D. A type of recurrent neural network for sequential processing

Answer: C

Q635. What is attention in sequence-to-sequence models?

A. A data augmentation method for creating synthetic text training data
B. A mechanism allowing the decoder to focus on relevant parts of the input sequence
C. A cross-entropy loss function for measuring sequence prediction accuracy
D. A regularization technique for preventing overfitting in sequence models

Answer: B

Q636. What is text embedding?

A. Converting text into dense numerical vectors
B. Removing text from the raw dataset
C. Encrypting text for secure storage
D. Converting numbers into readable text

Answer: A

Q637. What is part-of-speech (POS) tagging?

A. Removing individual words from the input text sequence
B. Translating text between two different natural languages
C. Assigning grammatical categories (noun, verb, etc.) to each word
D. Counting the number of syllables in each word

Answer: C

Q638. What is sequence-to-sequence (Seq2Seq) modeling?

A. Mapping a single input word to a single corresponding output word
B. Sorting input sequences into a predefined correct order
C. Clustering multiple sequences into meaningful related groups
D. Mapping an input sequence to an output sequence of potentially different length

Answer: C

Q639. What is cosine similarity used for in NLP?

A. Counting the total frequency of words in text
B. Parsing sentences into syntax tree structures
C. Generating new text from a trained model
D. Measuring the similarity between two text vectors

Answer: D

Q640. What is a language model?

A. A static vocabulary dictionary for word lookup
B. A rule-based grammar checker for syntax only
C. A database of aligned translation memory pairs
D. A model that predicts the probability of a sequence of words

Answer: D

Q641. What is the difference between semantic and instance segmentation?

A. Instance segmentation completely ignores the object class labels
B. Semantic segmentation is always the better approach overall
C. Semantic labels every pixel by class; instance distinguishes individual objects of the same class
D. Semantic and instance segmentation are completely identical approaches

Answer: B

Q642. What is data augmentation in computer vision?

A. Deleting images from the dataset to make it smaller and more manageable
B. Reducing the overall image quality to speed up the training process
C. Applying transformations like rotation, flipping, and cropping to increase training data
D. Converting all images to grayscale only without any other transformation

Answer: D

Q643. What is the YOLO algorithm?

A. A pixel-level semantic image segmentation algorithm
B. A supervised multi-class image classification algorithm
C. You Only Look Once - a real-time object detection algorithm
D. An unsupervised K-means image clustering algorithm

Answer: C

Q644. What is image feature extraction?

A. Adding textual annotation overlays onto the image pixels
B. Compressing images into smaller file archive formats
C. Identifying and representing distinctive visual patterns in images
D. Deleting specific features or layers from the image data

Answer: C

Q645. What is the purpose of max pooling?

A. Adding random noise to the feature map pixel values
B. Changing the color representation of the image data
C. Reducing spatial dimensions while retaining the most prominent features
D. Increasing the spatial size of the feature map dimensions

Answer: C

Q646. What is optical flow?

A. A type of spatial convolutional image filter kernel operation
B. A standard color model representation for digital images
C. The pattern of apparent motion of objects between consecutive frames
D. A container file format for storing compressed video data

Answer: C

Q647. What is the IoU (Intersection over Union) metric?

A. A metric measuring overlap between predicted and ground truth bounding boxes
B. A metric that measures overall image quality and clarity
C. A metric that assesses the overall color accuracy of images
D. A metric for evaluating the spatial resolution of images

Answer: C

Q648. What are anchor boxes in object detection?

A. The pixel coordinate system used for image representation
B. The individual color channels within the input image data
C. The border regions around the edges of an input image
D. Predefined bounding box shapes that help detect objects of various sizes and ratios

Answer: C

Q649. What is non-maximum suppression (NMS)?

A. A batch normalization method for stabilizing training
B. A post-processing step that removes overlapping duplicate detections
C. A gradient-based model training optimization technique
D. A non-linear activation function for hidden layers

Answer: B

Q650. What is the difference between object detection and image classification?

A. Image classification actually locates and draws bounding boxes around objects
B. Detection locates multiple objects with bounding boxes; classification assigns one label to the whole image
C. Object detection assigns a single label to the entire image without any localization
D. Object detection and image classification are completely identical tasks

Answer: B

Q651. What is stream processing?

A. Only storing data without processing it
B. Processing data in real-time as it arrives
C. Processing data in large scheduled batches
D. Only visualizing data without analysis

Answer: B

Q652. What is Apache Kafka?

A. A distributed event streaming platform for real-time data pipelines
B. A standalone relational database for transactional workloads
C. A supervised machine learning framework for model training
D. An interactive data visualization and charting dashboard

Answer: A

Q653. What is the difference between Spark and Hadoop MapReduce?

A. They are completely identical technologies with no performance differences
B. Spark only works effectively on small datasets under one gigabyte
C. MapReduce is significantly faster than Spark for all data workloads
D. Spark processes in-memory making it faster; MapReduce writes to disk between steps

Answer: D

Q654. What is NoSQL?

A. An interactive data visualization and charting dashboard
B. Non-relational databases designed for flexible schemas and horizontal scaling
C. A general-purpose compiled programming language for apps
D. A newer version of the standard SQL query language syntax

Answer: B

Q655. What is data partitioning?

A. Encrypting data at rest for privacy and compliance
B. Dividing data into smaller pieces distributed across multiple nodes
C. Merging all distributed data together into a single node
D. Deleting old data records to free up storage space

Answer: B

Q656. What is Apache Hive?

A. A distributed event streaming platform for real-time pipelines
B. A graph database for storing relationship network data
C. A data warehouse tool providing SQL-like queries over Hadoop data
D. A supervised machine learning library for model training

Answer: C

Q657. What is a Spark DataFrame?

A. A simple NumPy array structure
B. A distributed collection of data organized into named columns
C. A Python pandas DataFrame structure only
D. A traditional database table structure only

Answer: A

Q658. What is the CAP theorem?

A. It is primarily about effective data visualization techniques
B. It applies only to single standalone machines and servers
C. All three properties can always be fully achieved simultaneously
D. A distributed system can guarantee at most two of: Consistency, Availability, and Partition tolerance

Answer: B

Q659. What is data sharding?

A. Encrypting data at rest for privacy regulation compliance
B. Copying the entire database to every single server node
C. Splitting a database into smaller pieces across multiple servers
D. Deleting old data records to reclaim disk storage space

Answer: C

Q660. What is the role of a scheduler in big data systems?

A. Training supervised machine learning classification models
B. Visualizing data trends using interactive chart dashboards
C. Managing and coordinating the execution of distributed tasks
D. Storing and persisting data on distributed file systems

Answer: C

Q661. What is model drift?

A. Model training getting significantly faster with each iteration
B. Degradation in model performance over time due to changes in data
C. Model file size increasing automatically without configuration
D. Model accuracy consistently increasing without retraining

Answer: B

Q662. What is A/B testing in model deployment?

A. Comparing two model versions by serving them to different user groups
B. Running computations on two GPUs for faster processing speed
C. Training two different models simultaneously on the same data
D. Using two separate datasets for training and evaluation only

Answer: A

Q663. What is a feature store?

A. A model registry for tracking model versions and status
B. A code repository for managing source version control
C. A centralized repository for storing managing and serving ML features
D. A data warehouse for storing raw analytical data records

Answer: C

Q664. What is model serialization?

A. Deploying a model to a production serving endpoint
B. Saving a trained model to a file format that can be loaded later
C. Training a model on a labeled dataset from scratch
D. Evaluating a model's accuracy on a held-out test set

Answer: B

Q665. What is canary deployment?

A. Rolling back a deployed model to an earlier stable version
B. Gradually rolling out a new model to a small percentage of users first
C. Deploying the new model to all users immediately without rollout
D. Training a brand new model from scratch on latest data

Answer: B

Q666. What is the purpose of a model registry?

A. Storing and managing raw training data in a data lake
B. Centrally managing model versions metadata and deployment status
C. Visualizing model performance results on a dashboard
D. Writing and version-controlling source code for models

Answer: B

Q667. What is data drift?

A. Changes in the statistical distribution of input data over time
B. Changes in the source code of the training pipeline
C. Hardware failures in the production serving cluster
D. Changes in the internal model architecture configuration

Answer: A

Q668. What is the ONNX format?

A. A markup text file format for web page content
B. An open format for ML models enabling interoperability between frameworks
C. A proprietary relational database file storage format
D. A compressed video streaming container file format

Answer: B

Q669. What is a ML pipeline?

A. A database query for extracting records from tables
B. An automated sequence of steps from data processing to model deployment
C. An interactive data visualization charting dashboard
D. A single isolated model training step without automation

Answer: B

Q670. What is the role of Kubernetes in MLOps?

A. Visualizing model performance on dashboard charts
B. Orchestrating containerized ML services for scaling and management
C. Storing raw data files on a distributed file system
D. Training machine learning models on training data

Answer: B

Q671. What is the difference between individual and group fairness?

A. Group fairness ignores the outcomes of demographic groups
B. Individual fairness treats similar people similarly; group fairness ensures equal outcomes
C. They are completely identical fairness concepts with no differences
D. Individual fairness ignores the treatment of specific individuals

Answer: B

Q672. What is differential privacy?

A. A firewall system for blocking unauthorized network access
B. A process of permanently deleting all personal data records
C. A mathematical framework providing privacy guarantees when analyzing data
D. A standard encryption method for securing data files at rest

Answer: C

Q673. What is explainable AI (XAI)?

A. AI systems that require significantly less training data
B. AI systems that use significantly larger model architectures
C. AI systems that process data at significantly faster speeds
D. AI systems that provide interpretable explanations for their decisions

Answer: D

Q674. What is model interpretability?

A. The total file size in megabytes of the serialized model
B. The inference speed at which the model generates predictions
C. The overall accuracy metric of the trained classification model
D. The degree to which humans can understand a model's predictions

Answer: D

Q675. What is the black box problem in AI?

A. A hardware issue where the server casing is colored black
B. A storage problem where disk capacity runs out unexpectedly
C. Complex models making decisions that cannot be easily understood or explained
D. A networking error where connections time out intermittently

Answer: C

Q676. What are deepfakes?

A. Corrupted files that cannot be opened or processed
B. AI-generated synthetic media that convincingly replaces a person's likeness
C. Encrypted videos that require a decryption key to play
D. Low-quality images with significant compression artifacts

Answer: B

Q677. What is federated learning?

A. Sharing all raw training data openly across all participants
B. Training ML models across decentralized devices without sharing raw data
C. Training models exclusively on a single centralized server machine
D. Using only publicly available open-source datasets for training

Answer: B

Q678. What is the right to explanation under GDPR?

A. Individuals have the right to receive explanations for automated decisions
B. AI companies can keep their algorithms and models completely secret
C. No explanation is ever needed for any automated AI decisions
D. Only government agencies can request explanations of AI systems

Answer: A

Q679. What is data anonymization?

A. Making all personal data publicly available and fully open
B. Encrypting data with standard symmetric encryption only
C. Deleting all data from every storage system permanently
D. Removing or modifying personal identifiers so individuals cannot be identified

Answer: D

Q680. What is the trolley problem in AI ethics?

A. A moral dilemma about how autonomous systems should handle potential harm
B. A transportation route optimization problem for logistics planning
C. A data storage capacity issue for managing large files
D. A network routing problem for directing data packets

Answer: A

Q681. Which concept refers to an AI that can learn any intellectual task a human can?

A. Artificial Narrow Intelligence
B. Artificial Specific Intelligence
C. Artificial General Intelligence
D. Artificial Basic Intelligence

Answer: C

Q682. What distinguishes descriptive analytics from predictive analytics?

A. Descriptive forecasts future events while predictive summarizes past data
B. Descriptive needs real-time data while predictive works with batch data only
C. Descriptive uses deep learning while predictive relies on simple statistics
D. Descriptive summarizes past data while predictive forecasts future outcomes

Answer: D

Q683. In machine learning, what does the term 'generalization' mean?

A. Converting categorical variables into numerical representation
B. Memorizing every training sample exactly as it is provided
C. Performing well on new and previously unseen data points
D. Reducing the total number of features in a given dataset

Answer: C

Q684. Which of the following is an example of unsupervised learning?

A. Stock price prediction model
B. Customer segmentation clustering
C. Spam email classification
D. Handwriting digit recognition

Answer: B

Q685. What role does a loss function play in training a machine learning model?

A. It measures how far predictions deviate from actual values
B. It specifies the data format used for input preprocessing step
C. It determines the total memory required for model storage
D. It controls the number of features selected for training data

Answer: A

Q686. What is the difference between structured and unstructured data?

A. Structured data is always larger in file size than unstructured data is
B. Structured data requires GPUs while unstructured data uses only CPU cores
C. Structured data is only images while unstructured data is found in tables
D. Structured data uses rows and columns while unstructured data does not

Answer: D

Q687. Which concept describes a model that learns noise instead of true patterns?

A. Underfitting the training dataset
B. Overfitting the training data
C. Normalizing input features
D. Standardizing output labels

Answer: B

Q688. What is transfer learning in modern AI systems?

A. Transferring model weights from GPU memory to CPU memory
B. Converting one programming language into a different language
C. Moving data between two different storage systems for backup
D. Reusing a pre-trained model on a new but related task domain

Answer: D

Q689. Which metric is commonly used to evaluate classification model performance?

A. Mean Squared Error
B. Root Mean Absolute
C. Adjusted R-squared
D. Accuracy and F1-score

Answer: D

Q690. What is a 'feature' in machine learning terminology?

A. A visualization analysis tool
B. An individual measurable property
C. A type of network architecture
D. A bug found in software testing

Answer: B

Q691. What does the gradient of a function represent in optimization?

A. The maximum value the function can possibly achieve overall
B. The direction of steepest ascent at a given point in space
C. The total area under the curve of the function on a plot
D. The number of local minima in the function search space

Answer: B

Q692. What does Bayes' theorem allow us to compute?

A. The posterior probability given prior and likelihood
B. The determinant of any square matrix efficiently
C. The exact maximum value of a continuous function
D. The shortest path between nodes in a graph

Answer: A

Q693. What is the purpose of the chain rule in calculus for neural networks?

A. It selects which activation function to use in each layer
B. It determines the optimal number of hidden layers needed
C. It controls the learning rate during the training process
D. It enables gradient computation through composed functions

Answer: D

Q694. What does the covariance between two variables measure?

A. The exact ratio of one variable value to another variable
B. The causal relationship from one variable to the other one
C. The joint variability and direction of relationship between them
D. The maximum possible value either variable can take at all

Answer: C

Q695. What is the role of eigenvalues in Principal Component Analysis?

A. They determine the optimal number of training epochs to use
B. They specify the learning rate for gradient descent processes
C. They represent the variance amount captured by each component
D. They control the dropout probability in neural network layers

Answer: C

Q696. What property must a matrix have to be invertible?

A. All its elements must be positive real numbers
B. It must have more rows than total columns
C. Its determinant must be non-zero valued
D. All diagonal elements must equal exactly one

Answer: C

Q697. What does the standard deviation tell us about a dataset?

A. The spread of values around the mean value
B. The difference between maximum and minimum
C. The middle value when data is sorted in order
D. The most frequently occurring value in the data

Answer: A

Q698. In ML, what is a convex function?

A. A function that always produces negative output values only
B. A function with exactly two critical points at most always
C. A function only defined on integer input values overall
D. A function where any local minimum is also a global minimum

Answer: D

Q699. What does the L2 norm of a vector measure?

A. The Euclidean length or magnitude of vector
B. The largest absolute element in entire vector
C. The sum of all elements without absolute value
D. The count of non-zero elements in the vector

Answer: A

Q700. What is a probability distribution function?

A. A function describing likelihood of possible outcome values
B. A function that sorts data points in descending order value
C. A function removing outlier values from a given data set
D. A function converting categorical data into numerical codes

Answer: A

Q701. What is the difference between a shallow copy and a deep copy of a Python list?

A. Shallow copy works only with integers while deep copy handles all data types
B. Shallow copy duplicates all nested objects while deep copy copies references only
C. Shallow copy copies references to nested objects while deep copy duplicates all
D. Shallow copy is faster because it uses more memory than a deep copy overall

Answer: C

Q702. How does NumPy broadcasting work for arrays of different shapes?

A. It stretches smaller arrays to match larger ones following compatibility rules
B. It pads smaller arrays with zeros until both arrays have identical dimensions
C. It randomly samples elements from larger arrays to reduce them to smaller size
D. It raises an error whenever array shapes do not match exactly in all dimensions

Answer: A

Q703. What advantage does vectorized NumPy code have over standard Python loops?

A. Vectorized code uses more memory but produces more accurate results overall
B. Vectorized code executes in optimized C making it significantly much faster
C. Vectorized code only works with integer data types unlike Python loop code
D. Vectorized code automatically parallelizes across multiple machine clusters

Answer: B

Q704. What does the pandas groupby() method enable you to do?

A. Remove duplicate rows based on all column values in the frame
B. Sort DataFrame rows in ascending order by their index value names
C. Split data into groups and apply aggregate functions to each one
D. Merge two DataFrames horizontally using a common column as key

Answer: C

Q705. What is the purpose of Python's yield keyword in generator functions?

A. It terminates the function and returns the final value permanently
B. It creates a new thread for parallel function execution runs
C. It imports external modules required by function at call time
D. It pauses execution and produces a value without losing state

Answer: D

Q706. How do you create a virtual environment in Python for project isolation?

A. Using the command python -m venv env_name in the terminal window
B. Using the command pip create environment env_name from terminal
C. Using the command python --create-env env_name in the settings
D. Using the command python --isolate project env_name from shell

Answer: A

Q707. What is the role of the fit_transform() method in scikit-learn preprocessing?

A. It only transforms data using previously learned parameters from data
B. It fits preprocessor to data and transforms it in a single step
C. It evaluates model performance on the validation set after train
D. It splits the dataset into separate training and testing subsets

Answer: B

Q708. What does the lambda keyword create in Python?

A. A small anonymous function with a single expression body
B. A new class definition with automatic initialization method
C. A named function stored permanently in the global namespace
D. A decorator that modifies behavior of existing functions

Answer: A

Q709. What is the purpose of the pickle module in Python for ML workflows?

A. It provides functions for creating interactive data visualizations
B. It serializes Python objects to save and load trained ML models
C. It manages database connections for storing large training data
D. It handles HTTP requests for downloading datasets from the web

Answer: B

Q710. How does pandas merge() differ from concat() for combining DataFrames?

A. merge always produces fewer rows while concat always produces more rows
B. merge requires sorted data while concat works only with unsorted frames
C. merge joins on common columns while concat stacks along an axis direction
D. merge only works with numeric data while concat handles all data types

Answer: C

Q711. What is the difference between normalization and standardization?

A. Normalization works only with text while standardization works only with numeric data
B. Normalization removes outliers while standardization only handles missing values in data
C. Normalization scales to 0-1 range while standardization centers with zero mean unit variance
D. Normalization increases feature count while standardization decreases feature count total

Answer: C

Q712. Why might you use stratified sampling instead of simple random sampling?

A. It guarantees that the sample will be larger than simple random samples
B. It preserves the proportion of each class in the sample matching population
C. Stratified sampling is always faster than simple random sampling methods
D. Stratified sampling eliminates the need for any data cleaning steps after

Answer: B

Q713. What is the purpose of applying a log transformation to skewed data?

A. It converts categorical variables into numerical features for models
B. It reduces skewness making the distribution closer to normal shape
C. It removes all outliers automatically from the dataset completely
D. It increases the total number of samples available for training

Answer: B

Q714. When should you use label encoding versus one-hot encoding for categories?

A. Label encoding is used for numerical features while one-hot is for text only
B. One-hot encoding should be used only when the dataset has fewer than ten rows
C. Label encoding for ordinal categories and one-hot encoding for nominal ones
D. Label encoding is always superior to one-hot encoding in every scenario

Answer: C

Q715. What problem does the SMOTE technique address in preprocessing?

A. It handles missing values by predicting them using regression models
B. It removes highly correlated features from the dataset efficiently
C. It standardizes all features to have the same scale and distribution
D. It generates synthetic minority class samples to balance class distribution

Answer: D

Q716. What is the purpose of binning continuous variables into discrete groups?

A. To increase the total number of features in the dataset significantly
B. To convert numerical data into completely random categorical labels
C. To remove all missing values from the continuous variable columns
D. To reduce noise and capture non-linear patterns in the data values

Answer: D

Q717. How does the Interquartile Range method detect outliers in a dataset?

A. Values above the mean are always considered outliers by the IQR method
B. Only the single maximum and minimum values are considered outliers here
C. Values outside 1.5 times the IQR from Q1 and Q3 are flagged as outliers
D. Values within one standard deviation of median are flagged as outlier

Answer: C

Q718. What is data augmentation and when is it commonly used?

A. Encrypting sensitive data fields before storing them in the database
B. Removing features to reduce dimensionality in high-dimensional datasets
C. Converting unstructured data into structured tabular format for use
D. Creating modified copies of existing data to increase training set size

Answer: D

Q719. Why is it important to handle missing values before training a model?

A. Missing values always indicate that the entire dataset is unreliable here
B. Missing values only affect visualization and have no impact on models
C. Most ML algorithms cannot process datasets that contain any missing values
D. Handling missing values is optional since models automatically ignore them

Answer: C

Q720. What is feature scaling and why is it important for distance-based algorithms?

A. It converts all features to categorical type for better model interpretability
B. It duplicates important features to give them more weight during training
C. It ensures features contribute equally so larger-scale features do not dominate
D. It removes features with low variance to reduce the dimensionality of data

Answer: C

Q721. What does a right-skewed distribution indicate about the data?

A. Most values are clustered on the left with a long tail extending right
B. Most values are clustered on the right with a long tail extending left
C. The distribution has exactly two peaks at equal distances from center
D. Values are perfectly symmetrically distributed around the central mean

Answer: A

Q722. How does the Pearson correlation differ from the Spearman correlation?

A. Pearson requires sorted data while Spearman works with unsorted data only
B. Pearson only works with categorical data while Spearman works with numbers
C. Pearson measures linear relationships while Spearman measures monotonic ones
D. Pearson is always larger than Spearman for any given pair of variables

Answer: C

Q723. What is a pair plot used for in exploratory data analysis?

A. Visualizing the decision boundary of a trained classification model
B. Displaying all pairwise feature relationships and distributions simultaneously
C. Comparing model performance metrics across different algorithm choices
D. Showing only the relationship between the target variable and one feature

Answer: B

Q724. Why is it important to check for multicollinearity during EDA?

A. Checking for multicollinearity is optional and never affects final results
B. Multicollinearity only affects visualization and has no impact on any model
C. Multicollinearity always improves model accuracy and prediction performance
D. Highly correlated features can cause instability in regression model coefficients

Answer: D

Q725. What information does a Variance Inflation Factor provide?

A. The optimal number of clusters for a K-means clustering algorithm
B. The degree to which a feature is explained by other features in model
C. The total number of missing values in each feature of the dataset
D. The probability that a feature follows a normal distribution pattern

Answer: B

Q726. What is the purpose of a QQ plot in statistical analysis?

A. To compare model predictions against actual observed target values
B. To display the learning curve of a model during training process
C. To visualize the decision boundary of a classification model used
D. To assess whether data follows a specific theoretical distribution

Answer: D

Q727. How do violin plots improve upon standard box plots?

A. They only display the mean and are simpler to interpret than box plots
B. They show the actual probability density distribution shape of the data
C. They remove all outlier information to simplify the visualization look
D. They work only with categorical data unlike box plots which need numbers

Answer: B

Q728. What does the kurtosis of a distribution describe?

A. The asymmetry of the distribution around its mean value point
B. The tailedness and peakedness of the distribution compared to normal
C. The number of distinct modes or peaks present in the distribution
D. The total range between the minimum and maximum values observed

Answer: B

Q729. Why is it important to visualize data before applying machine learning algorithms?

A. Visualization is only for presentation and does not affect modeling choices
B. It is a mandatory step that all ML frameworks enforce before model training
C. Visualization automatically selects the best algorithm for the given dataset
D. It reveals patterns, outliers, and distributions that guide preprocessing decisions

Answer: D

Q730. What does a correlation matrix heatmap help identify?

A. The exact prediction accuracy of a trained machine learning model
B. Pairs of features with strong positive or negative linear relationships
C. The best train-test split ratio for the given dataset size overall
D. The optimal hyperparameters for a gradient boosting model training

Answer: B

Q731. How does regularization help prevent overfitting in linear regression models?

A. It adds a penalty term for large coefficients constraining model complexity
B. It increases the number of features to capture more complex data patterns
C. It multiplies all input features by a constant scaling factor for balance
D. It removes the least important training samples from the dataset entirely

Answer: A

Q732. What is the key assumption of Naive Bayes classification?

A. The decision boundary must be linear in the original feature space always
B. All features must be continuous numerical values with normal distribution
C. The dataset must have exactly equal numbers of samples in every class
D. Features are conditionally independent given the class label of the sample

Answer: D

Q733. How does the C parameter in SVM affect the decision boundary?

A. C determines the number of support vectors regardless of the data layout
B. Higher C creates a tighter margin penalizing misclassification more strictly
C. C only affects the kernel type used and has no impact on the margin width
D. Higher C creates a wider margin allowing more misclassification of points

Answer: B

Q734. What advantage does a decision tree have over logistic regression for certain tasks?

A. Decision trees always achieve higher accuracy than logistic regression models
B. Decision trees can capture non-linear relationships without feature engineering
C. Decision trees require less training data than logistic regression to converge
D. Decision trees never overfit regardless of the depth or number of features

Answer: B

Q735. What is the difference between hard and soft margin classification in SVM?

A. Hard margin is faster to train while soft margin requires exponentially more time
B. Hard margin uses linear kernels while soft margin always uses non-linear kernels
C. Hard margin allows no violations while soft margin permits some misclassification
D. Hard margin permits misclassification while soft margin requires perfect separation

Answer: C

Q736. Why might logistic regression outperform complex models on small datasets?

A. Its simplicity and fewer parameters make it less prone to overfitting on small data
B. Logistic regression always converges in fewer epochs than any other algorithm does
C. It automatically augments small datasets to increase the number of training samples
D. Logistic regression ignores the training data and uses only prior knowledge

Answer: A

Q737. What does the kernel trick enable SVM to do?

A. Map data to a higher-dimensional space to find non-linear decision boundaries
B. Reduce training time by eliminating the need for any optimization process
C. Convert regression problems into classification problems without data changes
D. Automatically select the best features from the dataset for classification

Answer: A

Q738. How does K in KNN affect the bias-variance tradeoff of the classifier?

A. Both small and large K always produce identical bias and variance values here
B. K has no effect on the bias-variance tradeoff of the KNN classifier at all
C. Small K has low bias and high variance while large K has high bias low variance
D. Small K has high bias and low variance while large K has low bias high variance

Answer: C

Q739. What is the purpose of the sigmoid function in logistic regression?

A. It normalizes input features to have zero mean and unit variance values
B. It maps raw output values to probabilities between zero and one range
C. It determines the optimal learning rate for gradient descent optimization
D. It selects the most important features for the model automatically here

Answer: B

Q740. Why is feature importance a useful output of tree-based models?

A. It guarantees that the model will achieve the highest possible accuracy
B. It automatically removes unimportant features from the training dataset
C. It converts categorical features into numerical features for processing
D. It reveals which features contribute most to the model prediction decisions

Answer: D

Q741. How does feature randomness in Random Forest improve model performance?

A. It guarantees each tree will have identical accuracy on test data set
B. It eliminates the need for any hyperparameter tuning in the model
C. It automatically removes all irrelevant features from the dataset used
D. It reduces correlation between trees leading to better ensemble diversity

Answer: D

Q742. What is the key difference between bagging and boosting ensemble techniques?

A. Bagging trains models sequentially while boosting trains all models in parallel
B. Bagging trains models in parallel independently while boosting trains sequentially
C. Bagging always outperforms boosting on every possible dataset and problem type
D. Bagging only works with neural networks while boosting only works with trees

Answer: B

Q743. What does the learning rate parameter control in gradient boosting?

A. The contribution of each tree to the final ensemble prediction output
B. The percentage of features randomly selected at each tree split
C. The number of trees to include in the final ensemble model overall
D. The maximum depth allowed for each individual decision tree used

Answer: A

Q744. What is stacking in the context of ensemble methods?

A. Using a meta-learner to combine predictions from diverse base models
B. Removing the worst-performing model from the ensemble iteratively
C. Training the same model architecture on different random data subsets
D. Simply averaging the predictions from all models in the ensemble

Answer: A

Q745. Why might an ensemble of weak learners outperform a single strong learner?

A. Combining diverse weak learners reduces variance without increasing bias much
B. Weak learners always have lower bias than any single strong learner model
C. Weak learners are always faster to train than any single strong learner is
D. An ensemble of weak learners never overfits regardless of ensemble size used

Answer: A

Q746. How does AdaBoost assign weights to training samples?

A. All samples receive equal weights throughout the entire training process
B. Misclassified samples receive higher weights so subsequent models focus on them
C. Weights are randomly assigned at each iteration without considering any errors
D. Correctly classified samples receive higher weights for reinforcement purposes

Answer: B

Q747. What is out-of-bag error estimation in Random Forest?

A. Evaluating each tree on samples not included in its bootstrap training set
B. Using a separate test set that was manually created before model training
C. Measuring training error on the full dataset after all trees are built
D. Running cross-validation with exactly five folds on the entire dataset used

Answer: A

Q748. What is the purpose of subsampling in Stochastic Gradient Boosting?

A. To convert the boosting algorithm into a bagging algorithm automatically
B. To ensure every sample is used exactly once across all trees in ensemble
C. To increase training time by processing every sample multiple times each
D. To reduce overfitting and training time by using random data subsets per tree

Answer: D

Q749. How does the max_depth parameter affect individual trees in an ensemble?

A. Deeper trees always improve ensemble accuracy without any risk of overfitting
B. Max depth has no effect on individual tree complexity or ensemble performance
C. Deeper trees have lower bias and higher variance capturing more complex splits
D. Deeper trees have higher bias and lower variance across ensemble predictions

Answer: C

Q750. Why is diversity important among base learners in an ensemble?

A. Diversity ensures all models make exactly the same predictions on every input
B. Diversity is irrelevant since all models should be identical for best results
C. Diverse models make different errors that can cancel out when combined together
D. Diverse models always have individually higher accuracy than identical models

Answer: C

Q751. How does DBSCAN determine cluster membership compared to K-Means?

A. DBSCAN requires the number of clusters while K-Means determines it automatically
B. DBSCAN groups points by density connectivity while K-Means uses centroid distance
C. DBSCAN uses distance to centroids while K-Means uses density-based grouping
D. Both algorithms use identical methods to determine cluster membership always

Answer: B

Q752. What is the elbow method used for in K-Means clustering?

A. To decide the maximum number of iterations for the K-Means convergence
B. To determine the optimal number of features to use for the clustering task
C. To evaluate whether K-Means or DBSCAN is the better algorithm for the data
D. To select the optimal K by finding the point where inertia reduction slows

Answer: D

Q753. What does the silhouette score measure in clustering evaluation?

A. How well each point fits its cluster versus the nearest neighboring cluster
B. The accuracy of cluster labels compared to true class labels in dataset
C. The total computational time required to train the clustering algorithm
D. The number of outliers present in each cluster after model convergence

Answer: A

Q754. How does hierarchical clustering differ from K-Means in its approach?

A. Hierarchical builds a tree of nested clusters while K-Means uses flat partitions
B. Hierarchical requires specifying K while K-Means does not need any parameters
C. Hierarchical only works with text data while K-Means works with numbers only
D. Hierarchical always produces fewer clusters than K-Means on the same data set

Answer: A

Q755. What is the role of the epsilon parameter in DBSCAN?

A. It defines the minimum number of clusters the algorithm must discover
B. It sets the maximum distance for two points to be considered neighbors
C. It controls the learning rate for optimizing the cluster assignments
D. It specifies the dimensionality reduction ratio before clustering starts

Answer: B

Q756. Why might PCA lose important information during dimensionality reduction?

A. PCA always preserves all information regardless of components selected count
B. It projects onto directions of maximum variance and may discard informative ones
C. PCA can only reduce to exactly two dimensions and no other number of them
D. It requires labeled data and fails when applied to unsupervised datasets

Answer: B

Q757. What is the difference between agglomerative and divisive hierarchical clustering?

A. Agglomerative starts with individual points and merges while divisive starts whole and splits
B. Both approaches always produce identical clustering results on any dataset given to them
C. Agglomerative only works with numerical data while divisive handles categorical data only
D. Agglomerative starts with one cluster and splits while divisive starts with individual points

Answer: A

Q758. How does t-SNE differ from PCA for dimensionality reduction?

A. Both methods produce identical visualizations on any dataset given to them always
B. PCA is non-linear while t-SNE is a purely linear dimensionality reduction method
C. PCA preserves global linear structure while t-SNE preserves local non-linear structure
D. PCA can only reduce to two dimensions while t-SNE works with any target dimension

Answer: C

Q759. What challenge does the curse of dimensionality pose for clustering algorithms?

A. High-dimensional data always forms perfectly separated clusters automatically here
B. It makes clustering faster because more dimensions provide more information now
C. The curse of dimensionality only affects supervised learning and never clustering
D. In high dimensions distance metrics become less meaningful reducing cluster quality

Answer: D

Q760. What is Gaussian Mixture Model clustering based on?

A. A density-based approach identical to DBSCAN using epsilon neighborhoods
B. Probabilistic assignment assuming data comes from a mixture of Gaussian distributions
C. Hard assignment of each point to exactly one cluster using closest centroid
D. A hierarchical approach that builds a dendrogram of cluster relationships

Answer: B

Q761. Why is accuracy a poor metric for imbalanced classification problems?

A. Accuracy requires balanced classes and cannot be computed on imbalanced data sets
B. Accuracy is only defined for regression problems and not for classification tasks
C. A model predicting only the majority class achieves high accuracy despite being useless
D. Accuracy always gives a value of zero when classes are imbalanced in the dataset

Answer: C

Q762. What is the ROC curve and what does it visualize?

A. A plot of training loss versus validation loss over each training epoch number
B. A plot of precision versus recall for different classification threshold values
C. A plot of model complexity versus generalization error for model selection use
D. A plot of true positive rate versus false positive rate at different thresholds

Answer: D

Q763. How does K-fold cross-validation improve upon a simple train-test split?

A. It randomly assigns labels to the test set to simulate real-world data distribution
B. It always uses exactly two folds regardless of the K value specified by the user
C. It uses all data for both training and testing reducing variance of performance estimate
D. It eliminates the need for any test data by using only training data for evaluation

Answer: C

Q764. What is the difference between micro and macro averaging for multi-class metrics?

A. Micro averages per class then takes mean while macro aggregates all predictions globally
B. Micro aggregates predictions globally while macro averages per class metrics equally
C. Micro averaging is used for binary classification while macro is for multi-class only
D. Both micro and macro averaging always produce exactly identical metric values here

Answer: B

Q765. What does the Area Under the ROC Curve represent?

A. The probability a random positive sample is ranked higher than a random negative one
B. The exact accuracy of the model at the optimal classification threshold selected
C. The maximum number of features the model can effectively use for prediction
D. The total training time required to achieve the best validation set performance

Answer: A

Q766. When should you use stratified K-fold cross-validation instead of standard K-fold?

A. When class distributions are imbalanced to ensure each fold represents all classes
B. When all features are categorical and need special handling during evaluation
C. When the dataset is very large and standard K-fold is computationally too expensive
D. When the dataset has perfectly balanced class proportions across all groups

Answer: A

Q767. What is the bias-variance tradeoff in model evaluation?

A. Bias and variance are independent and changing one has no effect on the other one
B. Higher bias always leads to higher variance and both increase simultaneously always
C. The tradeoff only applies to neural networks and not to any traditional ML algorithms
D. Simpler models have high bias low variance while complex models have low bias high variance

Answer: D

Q768. What is the purpose of a learning curve in model evaluation?

A. To visualize the feature importance rankings from a trained random forest model
B. To display the optimal hyperparameters found during grid search tuning process
C. To show how training and validation performance change with training set size
D. To plot the distribution of prediction errors across all test set observations

Answer: C

Q769. How does the log loss metric differ from accuracy for evaluation?

A. Log loss and accuracy always produce identical rankings of model performance
B. Log loss can only be used with regression models while accuracy is for classifiers
C. Log loss measures prediction speed while accuracy measures correctness of results
D. Log loss penalizes confident wrong predictions more heavily than accuracy does

Answer: D

Q770. What is the purpose of a calibration curve in model evaluation?

A. To assess whether predicted probabilities match actual observed frequencies
B. To determine the optimal number of features to include in the final model
C. To measure how fast the model makes predictions on new test data samples
D. To visualize the convergence of the training loss during model optimization

Answer: A

Q771. What is the wrapper method for feature selection?

A. Selecting features alphabetically by their column name in the original data frame
B. Using a model to evaluate different feature subsets and selecting the best one found
C. Selecting features based only on their individual correlation with the target variable
D. Removing all features with missing values regardless of their predictive importance

Answer: B

Q772. How does target encoding work for high-cardinality categorical features?

A. It replaces each category with the mean of the target variable for that group
B. It replaces each category with a randomly assigned numerical identifier value
C. It creates one binary column for each unique category in the feature column
D. It removes all categories that appear fewer than ten times in the dataset

Answer: A

Q773. What is the purpose of creating interaction features between variables?

A. To convert all numerical features into categorical features for tree models
B. To reduce the total number of features in the dataset to save memory space
C. To capture combined effects that individual features alone cannot represent
D. To remove multicollinearity between features by merging them into one column

Answer: C

Q774. Why is the filter method for feature selection computationally efficient?

A. It requires training the model multiple times with different feature subsets each
B. It evaluates features independently of any model using statistical measures only
C. It uses a complex model to evaluate every possible subset of feature combinations
D. It only works with datasets that have fewer than one hundred features total

Answer: B

Q775. What is the benefit of creating date-based features from timestamps?

A. Machine learning algorithms automatically extract date features without engineering
B. Date-based features always decrease model accuracy compared to raw timestamps
C. Extracting components like day of week and month reveals cyclical patterns in data
D. Timestamps are already in the optimal format for all machine learning algorithms

Answer: C

Q776. How does the Variance Threshold method work for feature selection?

A. It increases the variance of low-variance features through data augmentation
B. It selects features with the highest variance above the target variable mean
C. It calculates the variance between features to detect multicollinearity issues
D. It removes features with variance below a specified threshold as uninformative

Answer: D

Q777. What is feature hashing and when is it useful?

A. It sorts features by importance and removes the least important ones from data
B. It encrypts feature values for security during model training in production use
C. It maps high-dimensional categorical features to fixed-size vectors using a hash
D. It creates polynomial features from the original feature set for better models

Answer: C

Q778. Why might you apply a log transformation to a highly skewed feature?

A. It removes all outliers from the feature by compressing them to zero values
B. Log transformation converts categorical features to numerical features directly
C. Log transformation always improves model accuracy regardless of distribution shape
D. It reduces skewness and helps satisfy normality assumptions of some algorithms

Answer: D

Q779. What is the embedded method for feature selection?

A. Selecting features by running the model on every possible subset combination
B. Selecting features before training any model using only statistical measures
C. Feature selection performed as part of the model training process itself directly
D. Removing all features except the one with the highest correlation to target

Answer: C

Q780. How do you handle cyclical features like hour of day or day of week?

A. Treat them as regular continuous numerical features without any transformation
B. Simply remove cyclical features since they do not contain useful information
C. Encode them using sine and cosine transformations to preserve cyclic continuity
D. Convert them to categorical features and apply one-hot encoding to each value

Answer: C

Q781. Why does the vanishing gradient problem occur in deep networks?

A. Because the learning rate is always set too high for deep architectures
B. The problem only occurs in shallow networks and never in deep architectures
C. Deep networks always have too few parameters to compute meaningful gradients
D. Gradients shrink exponentially through many layers making early layers learn slowly

Answer: D

Q782. How does batch normalization improve deep neural network training?

A. It normalizes layer inputs reducing internal covariate shift for faster training
B. It removes the need for any activation functions in the neural network layers
C. It replaces the optimizer and directly updates weights without gradient descent
D. It increases the batch size automatically to improve training speed for models

Answer: A

Q783. What is the purpose of dropout regularization in neural networks?

A. It increases the number of neurons in each layer to improve model capacity
B. It drops entire training samples that the model already classifies correctly
C. It randomly deactivates neurons during training to prevent co-adaptation of features
D. It permanently removes neurons that have zero weight values in the network

Answer: C

Q784. How does the Adam optimizer combine momentum and adaptive learning rates?

A. Adam uses only momentum without any adaptive learning rate adjustment at all
B. It maintains running averages of both gradients and squared gradients for updates
C. It uses a fixed learning rate that never changes throughout the training process
D. Adam ignores gradient history and uses only the current gradient for each update

Answer: B

Q785. What is the difference between a dense layer and a convolutional layer?

A. Dense layers connect every neuron while convolutional layers use local receptive fields
B. Dense layers only work with images while convolutional layers work with tabular data
C. Dense layers use local connections while convolutional layers connect every neuron pair
D. Both layer types are identical and can be used interchangeably in any architecture

Answer: A

Q786. Why is the learning rate one of the most important hyperparameters?

A. The learning rate only affects the final accuracy and not the training speed at all
B. It only matters for the first epoch and is automatically adjusted after that point
C. The learning rate has no effect on training dynamics and can be set to any value
D. Too high causes divergence and too low causes extremely slow convergence of loss

Answer: D

Q787. What is transfer learning in the context of deep learning?

A. Transferring data from one storage system to another for model training use
B. Using a pre-trained network and fine-tuning it for a different related task
C. Training a model from random initialization on a completely new dataset only
D. Converting a classification model into a regression model without retraining

Answer: B

Q788. How does gradient clipping prevent training instability in deep networks?

A. It caps gradient values at a maximum threshold preventing exploding gradients
B. It removes all gradients entirely preventing any weight updates from occurring
C. It increases gradient values when they are too small to accelerate convergence
D. It replaces all negative gradients with zero to ensure only positive updates

Answer: A

Q789. What is the purpose of an embedding layer in deep learning?

A. To reduce the learning rate during training to prevent overfitting of model
B. To map discrete tokens to dense continuous vector representations efficiently
C. To normalize the input data before it enters the first hidden layer values
D. To compute the final loss value at the output of the neural network model

Answer: B

Q790. Why do deep learning models typically require GPU acceleration?

A. GPUs have more storage space than CPUs for holding the training dataset files
B. GPUs excel at parallel matrix operations which dominate deep learning computation
C. GPUs are required by all deep learning frameworks and cannot use CPU at all
D. GPUs automatically optimize hyperparameters while CPUs cannot perform this task

Answer: B

Q791. How does the LSTM cell solve the vanishing gradient problem of standard RNNs?

A. It processes sequences in reverse order to reduce the effective sequence length
B. It uses a much larger learning rate than standard RNN implementations typically use
C. It uses gates to control information flow allowing gradients to persist long-term
D. It removes all activation functions to ensure gradients never shrink in magnitude

Answer: C

Q792. What is the purpose of the discriminator in a GAN?

A. To distinguish between real data samples and generated fake data samples
B. To preprocess input data before it is fed to the generator network model
C. To generate realistic synthetic data samples from random noise input vectors
D. To evaluate the final quality of the trained model on a held-out test set

Answer: A

Q793. How does dilated convolution differ from standard convolution in CNNs?

A. Dilated convolution uses smaller filters than standard convolution always does
B. Standard convolution has larger receptive fields than dilated convolution does
C. Dilated convolution inserts gaps between filter elements expanding receptive field
D. Dilated convolution only works with one-dimensional input data like text only

Answer: C

Q794. What is the key innovation of the self-attention mechanism in Transformers?

A. It eliminates the need for any training data by using pre-defined attention maps
B. It requires data to be processed in strict sequential order one step at a time
C. It only attends to the immediately preceding element in the input sequence
D. It computes relationships between all positions in a sequence simultaneously

Answer: D

Q795. What problem does mode collapse cause in GAN training?

A. The discriminator becomes too weak to provide any useful training signal at all
B. The generator produces diverse outputs but none of them look like real data
C. The generator produces realistic but limited variety of outputs ignoring modes
D. The training process converges too slowly requiring excessive computation time

Answer: C

Q796. How do 1x1 convolutions serve as a dimensionality reduction technique in deep networks?

A. They only work with one-dimensional data and cannot be applied to image features
B. They remove the need for any pooling layers in the convolutional neural network
C. They increase the spatial dimensions of the feature maps while reducing channels
D. They reduce the number of channels by combining feature maps with learned weights

Answer: D

Q797. What is the purpose of positional encoding in Transformer architectures?

A. It replaces the need for any attention mechanism in the Transformer model
B. It determines the optimal position for inserting new layers during training
C. It encodes the position of the model parameters in memory for faster access
D. It injects sequence position information since self-attention has no order notion

Answer: D

Q798. How does depthwise separable convolution reduce computational cost?

A. It removes all bias terms from the convolutional layers to reduce parameters
B. It skips every other spatial position during convolution to halve computation
C. It uses only the first channel of input and ignores all other channels entirely
D. It separates spatial and channel-wise filtering reducing multiply-add operations

Answer: D

Q799. What is the role of the encoder and decoder in a sequence-to-sequence model?

A. The encoder handles training while the decoder is only used during evaluation
B. The encoder generates output and decoder processes the raw input sequence data
C. Both encoder and decoder perform identical operations on the same input data set
D. The encoder compresses input to a representation and decoder generates the output

Answer: D

Q800. What advantage does the GRU have over the standard LSTM architecture?

A. GRU uses fewer gates and parameters achieving similar performance more efficiently
B. GRU can process images while LSTM is limited to only text and audio data types
C. GRU does not use any gating mechanism unlike LSTM which uses three gates total
D. GRU always achieves higher accuracy than LSTM on every sequential task given

Answer: A

Q801. How does TF-IDF improve upon simple bag-of-words representation?

A. It counts words more quickly than bag-of-words without changing the representation
B. It weights words by importance reducing the influence of common uninformative terms
C. It only considers the first and last words in each document ignoring all others
D. TF-IDF and bag-of-words produce identical representations for all text documents

Answer: B

Q802. What is the key innovation of Word2Vec compared to one-hot encoding?

A. Word2Vec only works with very small vocabularies of less than one hundred words total
B. Word2Vec produces identical representations as one-hot encoding for all vocabulary items
C. Word2Vec requires labeled data for every word pair relationship in the corpus used
D. Word2Vec creates dense vectors capturing semantic relationships between words efficiently

Answer: D

Q803. What is the difference between stemming and lemmatization?

A. Stemming applies crude rules while lemmatization uses vocabulary and morphological analysis
B. Stemming uses dictionary-based root finding while lemmatization uses rule-based cutting
C. Stemming is more accurate than lemmatization for all natural language processing tasks
D. Both methods always produce identical output for every word in all languages used

Answer: A

Q804. How does the attention mechanism in seq2seq models improve machine translation?

A. It allows the decoder to focus on relevant source words for each output word generated
B. Attention removes the need for both encoder and decoder in the translation pipeline
C. It forces the decoder to translate each word independently without any context used
D. It limits the model to translating only between English and one other target language

Answer: A

Q805. What is the purpose of byte-pair encoding in modern NLP tokenizers?

A. It encrypts text using byte-level operations for secure communication purposes
B. It splits every word into individual characters without any merging of subwords
C. It converts all text to lowercase and removes all punctuation marks from strings
D. It iteratively merges frequent character pairs creating a subword vocabulary set

Answer: D

Q806. How does the BERT model differ from earlier language models in its training approach?

A. BERT uses bidirectional context by masking random tokens and predicting them
B. BERT uses left-to-right unidirectional context like all previous language models did
C. BERT processes text character by character unlike word-level previous models
D. BERT requires no pre-training and is trained only on downstream task data sets

Answer: A

Q807. What is the purpose of the CLS token in BERT?

A. It separates individual characters within each word of the input sequence
B. It indicates missing or unknown words in the vocabulary of the tokenizer
C. It provides a pooled representation for classification tasks using the model
D. It marks the end of every sentence in the input text sequence provided

Answer: C

Q808. What is perplexity and how is it used to evaluate language models?

A. It measures the vocabulary size of the language model's training corpus data
B. It measures the training time required to converge to the optimal solution found
C. It counts the number of parameters in the language model architecture structure
D. It measures how well the model predicts text with lower values being better scores

Answer: D

Q809. How does transfer learning with pre-trained models benefit NLP tasks with limited data?

A. Transfer learning only works when the new task has more data than the pre-training set
B. Pre-trained models must always be retrained from scratch on new task specific data
C. Pre-trained models provide rich language understanding that transfers to new tasks well
D. Pre-trained models can only be used for the exact task they were originally trained on

Answer: C

Q810. What is the difference between extractive and abstractive text summarization?

A. Extractive selects existing sentences while abstractive generates new sentences from scratch
B. Extractive only works with short texts while abstractive only works with long documents
C. Both methods always produce identical summaries for any given input document provided
D. Extractive generates new sentences while abstractive selects existing sentences from input

Answer: A

Q811. How does transfer learning with pre-trained CNNs benefit computer vision tasks with limited data?

A. Transfer learning only works for text classification and not for any image-based tasks
B. Early layers learn general features that transfer well reducing data needs for new tasks
C. Pre-trained models always perform worse than randomly initialized models on new tasks
D. Pre-trained models must be completely retrained from scratch for every new CV task

Answer: B

Q812. What is the difference between semantic and instance segmentation?

A. Semantic segmentation only works on grayscale images while instance works on color
B. Semantic labels each pixel by class while instance distinguishes individual objects too
C. Both segmentation types always produce identical pixel-level output maps on images
D. Instance labels each pixel by class while semantic distinguishes individual objects only

Answer: B

Q813. How does the YOLO architecture achieve real-time object detection?

A. It only detects objects in the center of the image ignoring all border regions always
B. It frames detection as a single regression problem processing the entire image at once
C. It processes each potential object region separately using a sliding window approach
D. YOLO requires multiple passes through the image increasing detection time significantly

Answer: B

Q814. What is the purpose of non-maximum suppression in object detection?

A. To maximize the size of all bounding boxes to ensure objects are fully contained
B. To increase the number of detected bounding boxes for better detection coverage
C. To remove redundant overlapping bounding boxes keeping only the best predictions
D. To suppress all detections with confidence scores above a certain threshold level

Answer: C

Q815. How does image augmentation help prevent overfitting in computer vision models?

A. It creates diverse variations forcing the model to learn invariant features robustly
B. It reduces the training set size to prevent the model from memorizing all samples
C. Augmentation always causes more overfitting because it adds artificial noise to data
D. It only works with very large datasets and has no benefit for small training sets

Answer: A

Q816. What is the Intersection over Union metric used for in object detection?

A. Measuring the total number of pixels in the input image being processed by model
B. Determining the optimal learning rate for the detection model training run
C. Measuring the overlap between predicted and ground truth bounding box regions
D. Calculating the training time for the object detection model on GPU hardware

Answer: C

Q817. How do feature pyramid networks improve multi-scale object detection?

A. They only detect objects at a single fixed scale ignoring all other size variations
B. They reduce all objects to the same size before detection to simplify the algorithm
C. They build a pyramid of feature maps at different scales for detecting objects of various sizes
D. Feature pyramids increase detection time without improving accuracy at any scale value

Answer: C

Q818. What is the role of anchor boxes in modern object detection architectures?

A. Anchor boxes are only used during evaluation and have no effect during training
B. They provide predefined reference boxes of various sizes and ratios for detection
C. They replace the non-maximum suppression step with a simpler averaging method
D. They fix the position of all detected objects to the center of the input image always

Answer: B

Q819. How does a U-Net architecture achieve precise segmentation of medical images?

A. It uses skip connections between encoder and decoder to preserve spatial details
B. U-Net only works with natural images and cannot process medical imaging data well
C. It requires pixel-level annotations for only ten percent of the training images used
D. It only uses downsampling layers without any upsampling to produce segmentation maps

Answer: A

Q820. What is the purpose of batch normalization in deep CNN architectures?

A. It normalizes layer activations stabilizing training and allowing higher learning rates
B. Batch normalization increases training time without providing any accuracy benefit
C. It only works with fully connected layers and cannot be used in convolutional layers
D. It removes all color information from input images before processing them further

Answer: A

Q821. How does Apache Spark improve upon Hadoop MapReduce for iterative ML algorithms?

A. Spark only works with small datasets while MapReduce handles large-scale data
B. Spark is always slower than MapReduce but provides better fault tolerance here
C. Spark writes intermediate results to disk just like MapReduce for reliability
D. Spark keeps data in-memory across iterations avoiding costly disk read-write cycles

Answer: D

Q822. What is the difference between batch processing and stream processing?

A. Both approaches process data identically and the terms are completely interchangeable here
B. Batch processing is always faster than stream processing for all data processing workloads
C. Batch processes data one item at a time while streaming processes all data at once together
D. Batch processes all data at once while streaming processes data as it arrives continuously

Answer: D

Q823. What is data sharding and why is it used in distributed databases?

A. It encrypts data using a shared key distributed across all machines in cluster
B. It horizontally partitions data across machines enabling scalability and performance
C. It compresses data into smaller files to reduce storage costs on each machine
D. It creates backup copies of all data on a single machine for redundancy purposes

Answer: B

Q824. How does Apache Spark SQL enable big data analytics for users familiar with SQL?

A. It converts all SQL queries into Python code before executing them on the cluster
B. It replaces Apache Spark entirely with a traditional database management system
C. Spark SQL only works with small datasets that fit in a single machine memory
D. It provides a SQL interface over distributed DataFrames for familiar query syntax

Answer: D

Q825. What is the CAP theorem and why is it relevant to distributed data systems?

A. It states distributed systems can have consistency, availability, and partition tolerance all at once
B. CAP theorem only applies to single-machine databases and not to distributed systems at all
C. It guarantees all distributed databases will always be faster than centralized database systems
D. It states only two of consistency, availability, and partition tolerance can be guaranteed together

Answer: D

Q826. What role does Apache Airflow play in big data engineering workflows?

A. It orchestrates and schedules complex data pipeline workflows as directed acyclic graphs
B. It trains machine learning models on big data using distributed gradient descent
C. It provides real-time stream processing capabilities for high-velocity data sources
D. It stores large datasets in distributed file systems across multiple cluster nodes

Answer: A

Q827. How does data replication improve fault tolerance in distributed systems?

A. Replication always degrades read performance to improve write speed of the system
B. It compresses data before storing to reduce the total storage space required
C. It copies data across multiple nodes so the system survives individual node failures
D. Replication stores data on a single machine to reduce network overhead costs

Answer: C

Q828. What is the difference between a data lake and a data warehouse?

A. Data lakes only support SQL queries while warehouses support all query languages used
B. Data lakes store raw multi-format data while warehouses store structured processed data
C. Data lakes store structured processed data while warehouses store raw unprocessed data
D. Both store data identically and the terms are completely interchangeable in practice

Answer: B

Q829. How does columnar storage format benefit analytical queries on big data?

A. Columnar storage is slower than row storage for all types of queries without exception
B. It stores data row by row which is optimal for reading entire records efficiently
C. It stores data column by column enabling efficient reads of only needed columns quickly
D. It only works with numerical data and cannot handle text or categorical columns well

Answer: C

Q830. What is the role of a cluster manager like YARN in distributed computing?

A. It converts distributed queries into single-machine queries for faster processing
B. It allocates resources and schedules tasks across nodes in the computing cluster
C. It handles data encryption and security for all files in the distributed system
D. It manages the final output visualization of processed big data results on screen

Answer: B

Q831. What is model drift and why does it require monitoring?

A. Model drift means the model is getting faster over time needing no intervention at all
B. Model performance degrades over time as data distributions change requiring retraining
C. Model drift only affects models deployed for less than one week in production systems
D. It refers to the physical movement of servers causing hardware failures in the system

Answer: B

Q832. How does canary deployment reduce risk when releasing new ML models?

A. It deploys the new model to all users simultaneously without any gradual rollout plan
B. It routes a small percentage of traffic to the new model while monitoring for issues
C. It only works for web applications and cannot be used with machine learning models
D. Canary deployment prevents any model from being deployed to production at all ever

Answer: B

Q833. What is the purpose of model serialization in deployment?

A. It deletes model weights that are no longer needed after training is completed
B. It visualizes the model architecture as a diagram for documentation purposes only
C. It converts trained models into a portable format for loading in production systems
D. It trains multiple models in sequence one after another on the same training data

Answer: C

Q834. How does infrastructure as code benefit MLOps workflows?

A. It eliminates the need for any cloud computing resources for ML model deployment
B. It requires all infrastructure to be manually configured through web interfaces only
C. It defines infrastructure in version-controlled code enabling reproducible environments
D. Infrastructure as code only applies to web development and not to ML workflows

Answer: C

Q835. What is the role of experiment tracking in the ML development lifecycle?

A. Experiment tracking is optional and provides no value to the ML development process
B. It logs parameters, metrics, and artifacts enabling comparison and reproducibility
C. It only tracks the final model and ignores all intermediate experiments and results
D. It automatically selects the best model without any human review or intervention

Answer: B

Q836. What is the shadow deployment pattern for ML models?

A. Shadow deployment completely ignores the old model predictions for all decisions
B. The new model replaces the old model immediately without any parallel operation
C. It deploys models only during nighttime hours when user traffic is at its lowest
D. The new model runs alongside the old one comparing outputs without serving users

Answer: D

Q837. How does model compression benefit edge deployment scenarios?

A. Model compression always improves accuracy while reducing the model size significantly
B. It reduces model size and latency enabling deployment on resource-constrained devices
C. Compression eliminates the need for any preprocessing of input data at inference
D. It only works with linear models and cannot be applied to deep neural networks

Answer: B

Q838. What is the purpose of data validation in an ML pipeline?

A. Data validation is unnecessary since all incoming data is always perfectly clean
B. To verify data quality and schema consistency before model training or inference
C. To train the model faster by removing data that takes long to process in pipeline
D. To convert all data types to strings for uniform processing in the ML pipeline

Answer: B

Q839. How does blue-green deployment work for ML model updates?

A. It trains the model in the blue environment and evaluates in the green one only
B. It gradually increases traffic to the new model over several weeks of monitoring
C. It maintains two identical environments and switches traffic between them instantly
D. Blue-green deployment only works with batch processing and not real-time serving

Answer: C

Q840. What is the benefit of using ONNX format for ML model interoperability?

A. ONNX always increases model inference time compared to native framework formats
B. ONNX only supports models trained with PyTorch and no other ML framework at all
C. ONNX provides a standard format allowing models to run across different frameworks
D. ONNX can only represent simple linear models and not complex neural networks

Answer: C

Q841. How can training data bias lead to discriminatory AI outcomes?

A. Training data bias always improves model performance and never causes discrimination
B. Bias in training data is automatically detected and removed by all ML algorithms used
C. If training data reflects historical discrimination the model learns and perpetuates it
D. Training data bias only affects model speed and has no impact on fairness of outputs

Answer: C

Q842. What is the difference between individual fairness and group fairness in AI?

A. Individual fairness focuses on groups while group fairness focuses on each individual person specifically
B. Both concepts are identical and always produce the same fairness outcomes for all situations
C. Individual fairness only applies to classification while group fairness only applies to regression
D. Individual fairness treats similar individuals similarly while group fairness ensures equal outcomes across groups

Answer: D

Q843. What is the right to explanation under GDPR for automated decision-making?

A. It requires companies to share all source code of their AI models publicly with everyone
B. Individuals have the right to obtain meaningful information about the logic of automated decisions
C. The right to explanation only applies to decisions made by human workers not AI systems
D. GDPR does not address automated decision-making at all and has no such requirement

Answer: B

Q844. How does differential privacy protect individual data in ML training?

A. Differential privacy removes all privacy protections to improve model accuracy scores
B. It adds calibrated noise to queries or gradients making individual data points unidentifiable
C. It requires all training data to be publicly available for anyone to inspect at any time
D. It encrypts the entire training dataset using standard encryption algorithms for security

Answer: B

Q845. What is the problem of automation bias in AI-assisted decision-making?

A. It refers to the bias introduced by automating the data collection process for ML
B. Automation bias only occurs in fully autonomous systems and never in AI-assisted ones
C. Users tend to over-rely on AI recommendations even when the AI may be incorrect here
D. Users never trust AI recommendations and always override them with their own judgment

Answer: C

Q846. How can AI systems inadvertently create feedback loops that amplify bias?

A. AI feedback loops always reduce bias over time through self-correcting learning loops
B. Biased predictions influence future data collection which further reinforces the bias
C. Feedback loops only occur in recommendation systems and never in other AI applications
D. AI systems never create feedback loops because they process each prediction independently

Answer: B

Q847. What is the dual-use dilemma in AI technology development?

A. AI technology can only be used for beneficial purposes and never for harmful applications
B. Dual-use only refers to AI running on two different hardware platforms simultaneously here
C. AI technology developed for harmful purposes can never be repurposed for beneficial uses
D. The same AI technology can be used for both beneficial and harmful purposes creating tension

Answer: D

Q848. What role does model interpretability play in building trust with AI stakeholders?

A. Model interpretability is only relevant for developers and has no value for end users
B. Interpretability reduces model accuracy so it should be avoided for all production models
C. All AI models are inherently interpretable and no additional effort is needed for trust
D. It enables stakeholders to understand and verify AI reasoning building confidence in system

Answer: D

Q849. How does the concept of AI safety differ from AI ethics?

A. AI safety is relevant only for AGI systems while AI ethics applies only to narrow AI systems
B. AI safety focuses on preventing AI from causing harm while ethics addresses broader societal values
C. AI safety and AI ethics are identical concepts with no meaningful difference between them
D. AI safety only concerns physical robots while AI ethics only concerns software applications

Answer: B

Q850. What is the challenge of measuring fairness across multiple protected attributes simultaneously?

A. Intersectional groups create complexity where satisfying fairness for one may violate another
B. Only one protected attribute can be considered at a time and others must be fully ignored
C. Multiple protected attributes have no interaction and can be handled completely independently
D. Fairness across multiple attributes is trivial to achieve by optimizing for single attribute

Answer: A

Q851. What is the difference between narrow AI and general AI?

A. Narrow AI is cheaper while general AI is expensive
B. Narrow AI handles specific tasks while general AI can handle any intellectual task
C. Narrow AI uses more data than general AI
D. There is no difference between them

Answer: B

Q852. What role does a reward signal play in reinforcement learning?

A. It provides labeled examples for classification
B. It tells the agent how good or bad its action was
C. It compresses the training data
D. It removes outliers from data

Answer: B

Q853. Which type of machine learning discovers hidden patterns in data without labels?

A. Supervised learning
B. Reinforcement learning
C. Unsupervised learning
D. Semi-supervised learning

Answer: C

Q854. What is the role of a validation set in the machine learning workflow?

A. To train the model parameters
B. To tune hyperparameters and prevent overfitting during development
C. To collect new data from users
D. To deploy the model to production

Answer: B

Q855. What is semi-supervised learning?

A. A method that uses only unlabeled data
B. A method that combines a small amount of labeled data with a large amount of unlabeled data
C. A method that only uses reinforcement signals
D. A method that does not use any data

Answer: B

Q856. What is the explore-exploit tradeoff in reinforcement learning?

A. Choosing between training and testing
B. Balancing trying new actions versus using known rewarding actions
C. Deciding between CPU and GPU usage
D. Selecting between classification and regression

Answer: B

Q857. What is a knowledge graph in the context of AI?

A. A chart showing training loss
B. A structured representation of real-world entities and their relationships
C. A neural network architecture
D. A type of decision tree

Answer: B

Q858. Why is data quality important for machine learning models?

A. Models always work regardless of data quality
B. Poor data quality leads to inaccurate or biased model predictions
C. Data quality only matters for image data
D. High quality data slows down training

Answer: B

Q859. What is an AI pipeline?

A. A physical pipe used in factories
B. A sequence of data processing and modeling steps from raw data to predictions
C. A single algorithm that solves all problems
D. A hardware component in computers

Answer: B

Q860. What is the difference between a parametric and non-parametric model?

A. Parametric models are always more accurate
B. Parametric models have a fixed number of parameters while non-parametric models grow with data
C. Non-parametric models cannot make predictions
D. There is no practical difference

Answer: B

Q861. What is the purpose of partial derivatives in machine learning?

A. To sort data alphabetically
B. To measure the rate of change of a function with respect to one variable while holding others constant
C. To create random numbers
D. To delete features from a dataset

Answer: B

Q862. What is a probability distribution?

A. A sorted list of numbers
B. A function that describes the likelihood of different outcomes for a random variable
C. A type of neural network
D. A data storage method

Answer: B

Q863. What is matrix multiplication and when is it defined?

A. Element-wise multiplication of any two matrices
B. Multiplying rows of the first matrix by columns of the second, defined when the first matrix's columns equal the second's rows
C. Always defined for any two matrices
D. Adding corresponding elements of two matrices

Answer: B

Q864. What is the normal distribution and why is it important in ML?

A. A distribution where all values are equal
B. A bell-shaped symmetric distribution that naturally arises in many real-world phenomena due to the central limit theorem
C. A uniform distribution over integers
D. A distribution only used in physics

Answer: B

Q865. What is the inverse of a matrix used for?

A. Transposing the matrix
B. Solving systems of linear equations and computing certain ML closed-form solutions
C. Increasing matrix dimensions
D. Sorting matrix elements

Answer: B

Q866. What is the central limit theorem?

A. A theorem about the center of a dataset
B. The sampling distribution of the mean approaches a normal distribution as sample size increases, regardless of the population distribution
C. All distributions are normal
D. Larger samples are always biased

Answer: B

Q867. What is the Euclidean distance between two points?

A. The Manhattan distance
B. The straight-line distance computed as the square root of the sum of squared differences
C. The cosine of the angle between them
D. The absolute difference of their means

Answer: B

Q868. What is the difference between a discrete and continuous random variable?

A. They are the same thing
B. A discrete variable takes countable values while a continuous variable can take any value in a range
C. Discrete variables are always larger
D. Continuous variables can only be integers

Answer: B

Q869. Why is the logarithm function commonly used in machine learning?

A. It makes all values negative
B. It converts multiplicative relationships to additive ones and helps with numerical stability
C. It is only used for display purposes
D. Logarithms are not used in ML

Answer: B

Q870. What is a correlation coefficient?

A. A measure of causation between variables
B. A value between -1 and 1 that measures the linear relationship strength between two variables
C. The mean of two variables
D. The number of data points

Answer: B

Q871. What is the difference between append() and extend() for Python lists?

A. They are identical
B. append() adds a single element while extend() adds all elements from an iterable individually
C. extend() adds a single element while append() adds multiple
D. Neither modifies the original list

Answer: B

Q872. What is the purpose of the with statement in Python?

A. Defining classes
B. Managing resources with automatic cleanup via context managers
C. Creating loops
D. Importing modules

Answer: B

Q873. How does pandas handle missing data by default?

A. It deletes the entire dataset
B. It represents missing values as NaN and provides methods like fillna() and dropna() to handle them
C. It replaces missing values with zeros automatically
D. It crashes when encountering missing data

Answer: B

Q874. What is slicing in NumPy arrays?

A. Deleting arrays
B. Extracting subarrays using index ranges with start:stop:step syntax
C. Sorting arrays
D. Concatenating arrays

Answer: B

Q875. What is the purpose of sklearn.preprocessing.StandardScaler?

A. It creates standard Python variables
B. It standardizes features by removing the mean and scaling to unit variance
C. It sorts data in standard order
D. It converts data to string format

Answer: B

Q876. What does the map() function do in Python?

A. Creates geographic maps
B. Applies a given function to each item of an iterable and returns the results
C. Maps keys to values in a dictionary
D. Measures memory allocation

Answer: B

Q877. How do you handle exceptions in Python?

A. Using if-else statements only
B. Using try-except blocks to catch and handle errors gracefully
C. Exceptions cannot be handled in Python
D. Using import statements

Answer: B

Q878. What is the difference between loc and iloc in pandas?

A. They are identical
B. loc uses label-based indexing while iloc uses integer position-based indexing
C. loc is faster than iloc always
D. iloc uses label-based indexing

Answer: B

Q879. What is a decorator in Python?

A. A way to add colors to output
B. A function that modifies the behavior of another function without changing its source code
C. A type of loop
D. A data visualization tool

Answer: B

Q880. What does the pandas pivot_table() method do?

A. Rotates a table physically
B. Creates a spreadsheet-style pivot table for summarizing data by grouping and aggregating
C. Deletes pivot columns
D. Transposes the DataFrame

Answer: B

Q881. What is the difference between min-max scaling and z-score standardization?

A. They produce identical results
B. Min-max scales to a fixed range like 0-1 while z-score centers data to zero mean and unit variance
C. Min-max is always better
D. Z-score only works with integers

Answer: B

Q882. What is multivariate imputation and when is it preferred?

A. Replacing missing values with zeros only
B. Using relationships between multiple features to estimate missing values, preferred when features are correlated
C. It only works for time series data
D. It is never preferred over simple imputation

Answer: B

Q883. Why is stratified sampling important when splitting imbalanced datasets?

A. It makes processing faster
B. It ensures each split maintains the same class distribution as the original dataset
C. It removes all minority class samples
D. It is only used for text data

Answer: B

Q884. What is the purpose of a pipeline in scikit-learn for preprocessing?

A. To create data visualizations
B. To chain multiple preprocessing steps and the model into a single object ensuring consistent application
C. To store data in databases
D. To generate synthetic data

Answer: B

Q885. What is the effect of class imbalance on model training?

A. It has no effect on model performance
B. Models tend to be biased toward the majority class, performing poorly on minority class predictions
C. It always improves accuracy
D. It only affects unsupervised learning

Answer: B

Q886. What is feature binarization?

A. Converting all features to text
B. Converting numerical features to binary values using a threshold
C. Deleting binary features
D. Doubling all feature values

Answer: B

Q887. How does the Robust Scaler differ from Standard Scaler?

A. They are identical
B. Robust Scaler uses median and IQR instead of mean and std, making it resistant to outliers
C. Robust Scaler is always slower
D. Standard Scaler handles outliers better

Answer: B

Q888. What is the purpose of winsorization in data preprocessing?

A. Deleting all outliers permanently
B. Capping extreme values at a specified percentile to reduce the impact of outliers
C. Converting data to Windows format
D. Increasing the number of outliers

Answer: B

Q889. Why should you encode categorical variables before feeding data to most ML algorithms?

A. Categorical data takes more storage
B. Most ML algorithms perform mathematical operations and cannot process text categories directly
C. Encoding is only needed for neural networks
D. Categorical variables are always irrelevant

Answer: B

Q890. What is the difference between MCAR, MAR, and MNAR missing data?

A. They are all the same type of missing data
B. MCAR is completely random, MAR depends on observed data, and MNAR depends on the missing values themselves
C. These are types of data encoding
D. They describe data augmentation methods

Answer: B

Q891. What is the difference between a kernel density plot and a histogram?

A. They are identical visualizations
B. A KDE provides a smooth continuous estimate of the probability density while a histogram uses discrete bins
C. KDE only works for categorical data
D. Histograms are always more accurate

Answer: B

Q892. How does a parallel coordinates plot help with multivariate EDA?

A. It only shows two variables at once
B. It visualizes each observation as a line across parallel axes for each feature, revealing clusters and patterns in high-dimensional data
C. It is a type of bar chart
D. It can only show categorical data

Answer: B

Q893. What does the chi-squared test assess in EDA?

A. The mean of a distribution
B. Whether there is a statistically significant association between two categorical variables
C. The slope of a regression line
D. The normality of data

Answer: B

Q894. Why is it useful to create a missing value heatmap during EDA?

A. To make the dataset look colorful
B. To visually identify patterns in missing data across features and observations
C. To remove all missing values automatically
D. To encrypt sensitive columns

Answer: B

Q895. What does a jointplot in Seaborn display?

A. Only a histogram
B. A bivariate plot showing the relationship between two variables along with their marginal distributions
C. A 3D surface plot
D. A pie chart with labels

Answer: B

Q896. What is the significance of the interquartile range in identifying outliers?

A. IQR identifies the mean value
B. Points falling below Q1 - 1.5*IQR or above Q3 + 1.5*IQR are considered potential outliers
C. IQR only applies to categorical data
D. IQR measures correlation

Answer: B

Q897. What is the purpose of a lag plot in time series EDA?

A. To create time delays in processing
B. To check for autocorrelation by plotting each observation against a lagged version of itself
C. To display GPS coordinates
D. To compress time series data

Answer: B

Q898. How does the Shapiro-Wilk test complement visual normality assessment?

A. It replaces all visualizations
B. It provides a formal statistical test for normality to supplement visual methods like QQ plots and histograms
C. It only tests for uniformity
D. It is identical to the chi-squared test

Answer: B

Q899. What does a swarm plot show that a box plot might miss?

A. The mean value
B. Individual data points showing the actual distribution density and potential clusters within groups
C. The standard deviation
D. Only outliers

Answer: B

Q900. What is the purpose of profiling tools like pandas-profiling during EDA?

A. To slow down analysis
B. To automatically generate comprehensive data quality reports including statistics, correlations, and missing value analysis
C. To convert data to PDF format
D. To deploy ML models

Answer: B

Q901. What is elastic net regularization and when is it preferred?

A. A type of neural network
B. A combination of L1 and L2 regularization that balances feature selection and coefficient shrinkage
C. A data preprocessing technique
D. A visualization method

Answer: B

Q902. How does a decision tree handle continuous features for splitting?

A. It ignores continuous features
B. It evaluates all possible threshold values to find the split that maximizes information gain or minimizes impurity
C. It converts them to binary first
D. It only uses the mean value

Answer: B

Q903. What is the difference between batch gradient descent and stochastic gradient descent?

A. They are identical algorithms
B. Batch uses the entire dataset per update while SGD uses one sample, trading accuracy per step for faster iterations
C. SGD uses the entire dataset
D. Batch gradient descent is always faster

Answer: B

Q904. Why is polynomial regression considered a linear model despite fitting curves?

A. It cannot fit curves
B. It is linear in its parameters even though it fits nonlinear relationships by using polynomial features of the input
C. It uses neural networks internally
D. Polynomial regression is not a linear model

Answer: B

Q905. What is the difference between one-vs-rest and one-vs-one for multiclass classification?

A. They produce identical results always
B. One-vs-rest trains K binary classifiers while one-vs-one trains K*(K-1)/2 classifiers for K classes
C. One-vs-rest only works for two classes
D. One-vs-one is always less accurate

Answer: B

Q906. What is the purpose of cost-sensitive learning?

A. To reduce the monetary cost of computing
B. To assign different penalties for different types of misclassification errors based on their real-world impact
C. To make models train faster
D. To reduce the number of features

Answer: B

Q907. How does early stopping work as a regularization technique?

A. It stops data collection early
B. It monitors validation performance during training and stops when performance starts degrading, preventing overfitting
C. It removes features early in the process
D. It only works for linear models

Answer: B

Q908. What is the role of maximum margin in SVM classification?

A. To minimize the margin between classes
B. To find the hyperplane that maximizes the distance between itself and the nearest data points from each class
C. To maximize training errors
D. To reduce the number of features

Answer: B

Q909. What is the difference between parametric and non-parametric supervised learning methods?

A. They are identical in approach
B. Parametric methods assume a fixed functional form while non-parametric methods make fewer assumptions and grow in complexity with data
C. Non-parametric methods always perform worse
D. Parametric methods have no parameters

Answer: B

Q910. What is the purpose of grid search in hyperparameter tuning?

A. To create data grids for visualization
B. To exhaustively search through a specified subset of hyperparameter combinations to find the best model configuration
C. To split data into grids
D. To remove outliers from grid-shaped data

Answer: B

Q911. How does the subsample parameter in gradient boosting affect the model?

A. It has no effect
B. Using a fraction of training data per tree introduces randomness that reduces overfitting similar to bagging
C. It increases training data size
D. It controls the number of features per split

Answer: B

Q912. What is the difference between soft and hard voting in ensemble classification?

A. They always produce identical results
B. Hard voting uses majority class votes while soft voting averages predicted probabilities before making the decision
C. Soft voting only works for regression
D. Hard voting uses probabilities

Answer: B

Q913. How does feature importance work in Random Forest?

A. It ranks features alphabetically
B. It measures how much each feature contributes to reducing impurity across all trees in the forest
C. It only considers the first tree
D. It is based solely on feature correlations

Answer: B

Q914. What is the role of the max_depth parameter in tree-based ensembles?

A. It controls the number of trees
B. It limits how deep each individual tree can grow, trading expressiveness for regularization
C. It sets the maximum number of features
D. It determines the learning rate

Answer: B

Q915. How does LightGBM achieve faster training than XGBoost?

A. It uses fewer trees
B. It uses histogram-based splitting and leaf-wise tree growth instead of level-wise, reducing computation
C. It skips the boosting step
D. It uses a simpler loss function

Answer: B

Q916. What is the purpose of regularization parameters like lambda and alpha in XGBoost?

A. To increase training speed only
B. To penalize complex trees by adding L1 and L2 penalties on leaf weights, reducing overfitting
C. To add more features
D. To skip cross-validation

Answer: B

Q917. What is blending in ensemble learning?

A. Mixing data from different datasets
B. A simplified stacking approach where base model predictions on a holdout set are used to train a meta-learner
C. Averaging all features together
D. Removing outliers from predictions

Answer: B

Q918. Why does boosting typically have higher bias than bagging at the start but lower at the end?

A. Boosting always has low bias
B. Boosting starts with a weak learner (high bias) and sequentially reduces bias by correcting errors, while bagging combines full models from the start
C. Bagging always has higher bias
D. Bias is irrelevant in ensemble methods

Answer: B

Q919. What is the purpose of column subsampling in gradient boosting?

A. To reduce the number of rows
B. To randomly select a subset of features for each tree, increasing diversity and reducing overfitting
C. To change the number of output classes
D. To modify the learning rate

Answer: B

Q920. How do you determine the optimal number of boosting rounds?

A. Always use exactly 100 rounds
B. Use early stopping on a validation set, stopping when validation performance stops improving
C. More rounds are always better
D. Use exactly as many rounds as features

Answer: B

Q921. What is the difference between K-Means and K-Medoids clustering?

A. They are identical algorithms
B. K-Means uses mean centroids while K-Medoids uses actual data points as cluster centers, making it more robust to outliers
C. K-Medoids is always faster
D. K-Means uses actual data points as centers

Answer: B

Q922. How does the explained variance ratio help in choosing the number of PCA components?

A. It is not useful for PCA
B. It shows the proportion of total variance captured by each component, helping choose enough components to retain a desired percentage of information
C. It measures clustering quality
D. It indicates the number of outliers

Answer: B

Q923. What is the difference between DBSCAN's epsilon and minPoints parameters?

A. They control the same thing
B. Epsilon defines the neighborhood radius while minPoints sets the minimum neighbors needed to form a dense core point
C. Epsilon counts points and minPoints measures distance
D. They are both optional parameters

Answer: B

Q924. What is the Calinski-Harabasz index for clustering evaluation?

A. A measure of classification accuracy
B. The ratio of between-cluster dispersion to within-cluster dispersion, where higher values indicate better-defined clusters
C. A feature selection metric
D. A dimensionality reduction technique

Answer: B

Q925. How does Gaussian Mixture Model clustering assign soft cluster memberships?

A. It uses hard assignments like K-Means
B. Each data point receives a probability of belonging to each cluster based on the Gaussian component likelihoods
C. It only assigns points to the nearest center
D. It does not produce cluster assignments

Answer: B

Q926. What is the difference between complete linkage and single linkage in hierarchical clustering?

A. They produce identical results
B. Complete linkage uses maximum inter-cluster distance while single linkage uses minimum, leading to different cluster shapes
C. Single linkage always produces better clusters
D. Complete linkage is single linkage with extra steps

Answer: B

Q927. What is the reconstruction error in autoencoders and why is it important?

A. It measures classification accuracy
B. It measures how well the autoencoder's output matches its input, indicating the quality of the learned representation
C. It counts the number of layers
D. It measures training speed

Answer: B

Q928. What is the role of the Expectation-Maximization algorithm in GMM?

A. It is a clustering evaluation metric
B. EM iteratively estimates cluster parameters (E-step: compute responsibilities, M-step: update parameters) until convergence
C. It only initializes clusters
D. It is used for supervised learning only

Answer: B

Q929. How does t-SNE preserve local structure during dimensionality reduction?

A. It preserves global structure instead
B. It converts pairwise distances to conditional probabilities and minimizes KL divergence between high and low-dimensional distributions
C. It uses linear transformations only
D. It discards all distance information

Answer: B

Q930. What is the purpose of the perplexity parameter in t-SNE?

A. It measures classification accuracy
B. It controls the effective number of nearest neighbors considered, balancing attention between local and global aspects of the data
C. It sets the output dimensions
D. It determines the number of clusters

Answer: B

Q931. What is the precision-recall tradeoff and why does it exist?

A. They always improve together
B. Increasing the decision threshold increases precision but decreases recall because fewer positive predictions are made but they are more confident
C. There is no tradeoff between them
D. They are the same metric

Answer: B

Q932. What is the advantage of using mean absolute percentage error for regression evaluation?

A. It is always the best metric
B. MAPE expresses errors as percentages making it scale-independent and intuitive for stakeholders to understand
C. It works well when actual values are near zero
D. MAPE is identical to RMSE

Answer: B

Q933. How does leave-one-out cross-validation work?

A. It removes one feature at a time
B. Each data point is used as a single test set while all remaining points are used for training, repeated for every point
C. It splits data into two halves
D. It only uses the first data point for testing

Answer: B

Q934. What is the adjusted R-squared and why is it better than regular R-squared?

A. They are identical metrics
B. Adjusted R-squared penalizes for the number of features, preventing artificial inflation from adding irrelevant variables
C. Adjusted R-squared is always lower
D. Regular R-squared accounts for model complexity

Answer: B

Q935. What is a classification report in scikit-learn?

A. A report about the dataset
B. A summary table showing precision, recall, F1-score, and support for each class along with averages
C. A list of all model parameters
D. A data quality report

Answer: B

Q936. When is the F-beta score more appropriate than the F1 score?

A. Never, F1 is always best
B. When you want to weight precision and recall differently, with beta > 1 favoring recall and beta < 1 favoring precision
C. When you have exactly two classes
D. When datasets are perfectly balanced

Answer: B

Q937. What is the purpose of a cumulative gains chart in model evaluation?

A. To show cumulative training loss
B. To show how much of the positive class is captured by targeting the top-scored percentage of predictions
C. To display feature importance
D. To measure computational gains

Answer: B

Q938. What is the difference between macro and weighted averaging for multi-class metrics?

A. They always produce the same result
B. Macro averaging treats all classes equally regardless of size while weighted averaging accounts for class frequency
C. Macro averaging is always preferred
D. Weighted averaging ignores class sizes

Answer: B

Q939. What is the Huber loss and when is it preferred over MSE?

A. They are identical losses
B. Huber loss is quadratic for small errors and linear for large errors, making it robust to outliers while still being differentiable
C. Huber loss is always worse than MSE
D. It only works for classification

Answer: B

Q940. What is the concept of an evaluation metric being proper?

A. It means the metric is correctly coded
B. A proper scoring rule is maximized when the predicted probabilities match the true probabilities, encouraging honest probability estimates
C. It means the metric is simple
D. Proper metrics are only for regression

Answer: B

Q941. How does frequency encoding work for categorical variables?

A. It converts categories to sound frequencies
B. It replaces each category with its frequency of occurrence in the dataset, capturing popularity information
C. It only works for time series
D. It is identical to one-hot encoding

Answer: B

Q942. What is the purpose of creating ratio and proportion features?

A. To normalize all features to 1
B. To capture relative relationships between variables that raw values alone may not express
C. To divide the dataset into portions
D. To calculate model accuracy

Answer: B

Q943. How do sine and cosine transformations encode cyclical features?

A. They convert cycles to straight lines
B. They map cyclical values to a circle using sin and cos, preserving the fact that values like 23:00 and 01:00 are close
C. They only work for audio data
D. They remove cyclical patterns from data

Answer: B

Q944. What is the difference between backward elimination and forward selection for feature selection?

A. They always select the same features
B. Backward elimination starts with all features and removes the least important, while forward selection starts empty and adds the most important
C. They are identical algorithms
D. Forward selection removes features

Answer: B

Q945. What is weight of evidence encoding for categorical variables?

A. It measures the physical weight of data
B. WoE measures the predictive power of each category by computing the log ratio of the distribution of events to non-events within each category
C. It only works for numerical features
D. It assigns random weights to categories

Answer: B

Q946. How does the correlation-based filter method work for feature selection?

A. It selects features randomly
B. It selects features that have high correlation with the target variable but low correlation with each other, reducing redundancy
C. It only removes constant features
D. It requires model training

Answer: B

Q947. What is the purpose of creating window-based features for time series data?

A. To resize data visualizations
B. To capture temporal patterns by computing statistics like rolling mean, sum, or standard deviation over fixed time windows
C. To remove time information
D. To encrypt temporal data

Answer: B

Q948. What is the Recursive Feature Elimination method?

A. A method that recursively adds features
B. RFE trains a model, removes the least important feature, retrains, and repeats until the desired number of features remains
C. It eliminates all features recursively
D. It is a type of data augmentation

Answer: B

Q949. Why might you create statistical aggregate features from grouped data?

A. To reduce the number of rows only
B. Aggregates like group means, counts, medians, and standard deviations capture entity-level behavioral patterns from transactional data
C. Aggregation always loses useful information
D. Statistical aggregates are only for visualization

Answer: B

Q950. What is the difference between one-hot encoding and binary encoding for high cardinality features?

A. They produce identical outputs
B. One-hot creates one column per category while binary encoding uses log2(n) columns with binary representation, greatly reducing dimensionality
C. Binary encoding creates more columns
D. One-hot always uses fewer columns

Answer: B

Q951. How does the Leaky ReLU activation function address the dying ReLU problem?

A. It has the same issue as ReLU
B. Instead of outputting zero for negative inputs, Leaky ReLU outputs a small negative slope, keeping neurons active and allowing gradient flow
C. It only works for positive inputs
D. It eliminates all negative values

Answer: B

Q952. What is the purpose of weight regularization in neural networks?

A. To make weights larger
B. To penalize large weight values in the loss function, constraining model complexity and reducing overfitting
C. To initialize weights to zero
D. To speed up inference

Answer: B

Q953. What is the difference between SGD with momentum and vanilla SGD?

A. They are identical
B. Momentum accumulates past gradients to smooth updates and accelerate convergence, especially in directions with consistent gradient sign
C. Vanilla SGD is faster
D. Momentum only works for small datasets

Answer: B

Q954. What is the Xavier/Glorot initialization strategy?

A. Setting all weights to one
B. Initializing weights from a distribution scaled by the number of input and output neurons to maintain signal variance across layers
C. Using random integers as weights
D. Copying weights from another model

Answer: B

Q955. How does data augmentation help in deep learning?

A. It reduces the dataset size
B. It artificially increases training data diversity by applying transformations, helping the model generalize and reducing overfitting
C. It cleans the data
D. It only works for text data

Answer: B

Q956. What is the difference between online learning and batch learning in neural networks?

A. They are the same approach
B. Online learning updates weights after each sample while batch learning processes the entire dataset before updating
C. Online learning is always better
D. Batch learning processes one sample at a time

Answer: B

Q957. What is the purpose of the He initialization for ReLU networks?

A. It is identical to Xavier initialization
B. He initialization scales weights by sqrt(2/fan_in) to account for ReLU's zeroing of negative values, maintaining proper signal variance
C. It initializes all weights to zero
D. It only works for sigmoid networks

Answer: B

Q958. What is label smoothing and how does it help training?

A. Making labels blurry in images
B. Replacing hard 0/1 labels with soft values like 0.1/0.9, reducing model overconfidence and improving generalization
C. It removes labels from the dataset
D. It only works for regression

Answer: B

Q959. What is the warmup strategy in learning rate scheduling?

A. Preheating the GPU hardware
B. Starting with a very small learning rate and gradually increasing it to the target rate over initial training steps
C. Using a constant high learning rate
D. Warmup is only used during inference

Answer: B

Q960. What is the purpose of gradient accumulation in deep learning?

A. To store gradients permanently
B. To simulate larger batch sizes by accumulating gradients over multiple mini-batches before updating weights, useful when GPU memory is limited
C. To remove gradients
D. It slows down training intentionally

Answer: B

Q961. How does the forget gate in an LSTM cell work?

A. It forgets the model parameters
B. It decides what information to discard from the cell state using a sigmoid gate that outputs values between 0 and 1
C. It removes training data
D. It deletes gradient values

Answer: B

Q962. What is the discriminator's role in GAN training?

A. It generates new data
B. It classifies inputs as real or fake, providing training signal to the generator to improve its outputs
C. It preprocesses data
D. It sets the learning rate

Answer: B

Q963. What is batch renormalization and why was it introduced?

A. It is identical to batch normalization
B. Batch renormalization corrects batch normalization's issues with small or non-iid mini-batches by using running statistics with trainable correction parameters
C. It removes normalization entirely
D. It only works for very large batches

Answer: B

Q964. What is the attention mechanism's key-query-value framework?

A. A database query system
B. Queries determine what to attend to, keys determine relevance scores, and values carry the information to be aggregated based on attention weights
C. Keys contain the input data only
D. Queries and keys are always identical

Answer: B

Q965. What is the purpose of the bottleneck layer in a ResNet block?

A. To slow down training
B. To reduce computational cost by using 1x1 convolutions to decrease and then increase the channel dimensions around the expensive 3x3 convolution
C. To create an autoencoder
D. To remove skip connections

Answer: B

Q966. How does teacher forcing work in training sequence generation models?

A. Using a teacher to label data
B. During training, the model receives the actual ground truth output token as input for the next step rather than its own predicted token
C. It only works during inference
D. It replaces the decoder entirely

Answer: B

Q967. What is group normalization and when is it preferred over batch normalization?

A. They are the same technique
B. Group normalization normalizes over groups of channels within each sample, independent of batch size, preferred when batch sizes are too small for stable batch statistics
C. Group norm requires large batches
D. It only works for RNNs

Answer: B

Q968. What is the difference between transposed convolution and upsampling in CNNs?

A. They are identical operations
B. Transposed convolution learns upsampling parameters while standard upsampling uses fixed interpolation methods like bilinear or nearest neighbor
C. Upsampling always produces better results
D. Transposed convolution reduces dimensions

Answer: B

Q969. What is the concept of feature reuse in DenseNet architecture?

A. Reusing training data
B. Each layer receives feature maps from all preceding layers as input, enabling direct feature reuse, gradient flow improvement, and parameter efficiency
C. Only the last layer reuses features
D. Feature reuse causes overfitting

Answer: B

Q970. What is knowledge distillation's dark knowledge?

A. Secret information hidden in models
B. The soft probability distributions over non-target classes produced by the teacher model, containing inter-class similarity information
C. Dark knowledge does not exist
D. It refers to unlabeled data

Answer: B

Q971. How does the Skip-gram model in Word2Vec work?

A. It skips words during processing
B. It predicts surrounding context words given a center word, learning word embeddings that capture semantic relationships
C. It classifies sentences into categories
D. It only works with consecutive words

Answer: B

Q972. What is the difference between GloVe and Word2Vec embeddings?

A. They produce identical embeddings
B. Word2Vec uses local context windows with neural networks while GloVe uses global word co-occurrence statistics with matrix factorization
C. GloVe is always better
D. Word2Vec uses global statistics

Answer: B

Q973. What is the masked language modeling objective used in BERT?

A. Removing all words from text
B. Randomly masking some input tokens and training the model to predict the masked tokens from bidirectional context
C. Translating masked text
D. Only predicting the next word

Answer: B

Q974. What is the role of the encoder in a machine translation system?

A. It translates the text directly
B. It processes the input sentence into a contextual representation that captures its meaning for the decoder to use
C. It only tokenizes the input
D. It generates the output translation

Answer: B

Q975. What is the difference between autoregressive and autoencoding language models?

A. They are identical approaches
B. Autoregressive models predict tokens left-to-right (GPT-style) while autoencoding models use bidirectional context with masked tokens (BERT-style)
C. Autoregressive models are always better
D. Autoencoding models generate text one token at a time

Answer: B

Q976. What is named entity recognition used for in practical applications?

A. Naming files on a computer
B. Identifying and classifying entities like person names, organizations, locations, and dates in text for information extraction
C. Renaming variables in code
D. Creating new entity types

Answer: B

Q977. How does beam search differ from greedy decoding for text generation?

A. They produce identical outputs
B. Beam search maintains multiple candidate sequences at each step, exploring a broader search space than greedy decoding which only keeps the single best token
C. Greedy decoding is always better
D. Beam search uses random sampling

Answer: B

Q978. What is the purpose of the attention mask in Transformer models?

A. To hide the model from users
B. To prevent the model from attending to padding tokens or future positions, ensuring valid attention computation
C. To mask training errors
D. To speed up computation without any functional purpose

Answer: B

Q979. What is text embedding and how does it differ from word embedding?

A. They are identical concepts
B. Text embedding produces a single vector for an entire sentence or document while word embedding produces individual vectors for each word
C. Text embedding only works for short texts
D. Word embedding handles entire documents

Answer: B

Q980. What is the purpose of the temperature parameter in language model generation?

A. It controls GPU temperature
B. It scales the logits before softmax, with lower values making output more focused and deterministic and higher values increasing randomness and diversity
C. It has no effect on output
D. It controls the learning rate

Answer: B

Q981. How does the region proposal network work in Faster R-CNN?

A. It manually selects regions
B. RPN slides a small network over the feature map to predict object proposals and objectness scores simultaneously
C. It uses random cropping for proposals
D. RPN is a separate model from the detector

Answer: B

Q982. What is the difference between top-1 and top-5 accuracy in image classification?

A. They measure the same thing
B. Top-1 requires the correct class to be the single highest prediction while top-5 counts it as correct if the true class is among the five highest predictions
C. Top-5 is always lower than top-1
D. They are used for different types of images

Answer: B

Q983. What is mean Average Precision in object detection evaluation?

A. The average image resolution
B. mAP averages the precision across different recall levels and across all object classes, providing a single performance metric for detection
C. It counts the number of detected objects
D. It measures processing speed

Answer: B

Q984. How does the U-Net skip connections help with image segmentation?

A. They skip the encoder entirely
B. Skip connections concatenate encoder features with decoder features, providing high-resolution spatial information that helps precise boundary localization
C. They remove low-level features
D. Skip connections only connect the first and last layers

Answer: B

Q985. What is the purpose of image augmentation techniques like random erasing?

A. To corrupt data permanently
B. Random erasing masks random rectangular patches in training images, forcing the model to rely on diverse features rather than specific local patterns
C. To reduce dataset size
D. To remove objects from images permanently

Answer: B

Q986. What is the difference between single-shot and multi-scale object detection?

A. They detect objects at the same scale
B. Multi-scale detection processes images at multiple resolutions or uses feature pyramids to detect objects of various sizes, while single-scale may miss small or large objects
C. Single-shot always detects more objects
D. Multi-scale only works for large objects

Answer: B

Q987. What is semantic segmentation's pixel-wise classification?

A. Classifying entire images into categories
B. Assigning a class label to every individual pixel in the image, creating a dense prediction map
C. Detecting objects with bounding boxes
D. Counting pixels in each color

Answer: B

Q988. How does Mixup data augmentation work for training CV models?

A. Mixing different datasets together
B. Creating new training samples by linearly interpolating between pairs of images and their labels, encouraging smoother decision boundaries
C. Mixing color channels randomly
D. Combining two models into one

Answer: B

Q989. What is the role of backbone networks in object detection architectures?

A. They provide structural support for servers
B. Backbone networks extract hierarchical visual features from the input image that are shared by detection head and other task-specific modules
C. They generate bounding boxes directly
D. Backbones only process text data

Answer: B

Q990. What is the difference between instance segmentation and panoptic segmentation?

A. They are identical tasks
B. Instance segmentation only handles countable objects while panoptic segmentation unifies both instance segmentation for things and semantic segmentation for stuff regions
C. Panoptic only detects backgrounds
D. Instance segmentation covers everything in the image

Answer: B

Q991. How does Apache Spark's lazy evaluation improve performance?

A. It makes Spark slower
B. Spark builds an execution plan without computing results until an action is called, enabling query optimization and avoiding unnecessary computations
C. It processes data immediately
D. Lazy evaluation has no effect on performance

Answer: B

Q992. What is the difference between a data lake and a data lakehouse?

A. They are identical concepts
B. A data lakehouse combines the flexibility of data lakes with the management and performance features of data warehouses, supporting both analytics and ML workloads
C. Data lakehouses only store structured data
D. Data lakes are always better

Answer: B

Q993. What is the purpose of data serialization formats like Parquet and Avro?

A. They are programming languages
B. They efficiently encode structured data for storage and transmission, with Parquet optimized for columnar analytics and Avro for row-based serialization
C. They only work with text data
D. They replace databases entirely

Answer: B

Q994. How does Apache Kafka enable real-time data streaming?

A. It stores data in batches only
B. Kafka provides a distributed publish-subscribe messaging system where producers publish events to topics and consumers process them in near real-time
C. It replaces all databases
D. Kafka only works with text data

Answer: B

Q995. What is the role of a distributed file system in big data?

A. It only stores small files
B. It stores large datasets across multiple machines, providing fault tolerance through replication and enabling parallel access to data
C. It replaces local file systems entirely
D. It only works with structured data

Answer: B

Q996. What is the purpose of a message queue in big data architectures?

A. To send emails between users
B. To decouple data producers from consumers, buffering messages and enabling asynchronous processing at different rates
C. To queue database queries
D. To sort messages alphabetically

Answer: B

Q997. How does data partitioning improve query performance in big data systems?

A. It always slows down queries
B. Partitioning divides data into segments based on key columns, allowing queries to scan only relevant partitions instead of the entire dataset
C. It creates duplicate data
D. Partitioning only works for small datasets

Answer: B

Q998. What is the difference between structured, semi-structured, and unstructured data in big data?

A. They are all the same format
B. Structured has fixed schema (tables), semi-structured has flexible schema (JSON/XML), and unstructured has no schema (images/text)
C. Only structured data is useful
D. Unstructured data cannot be processed

Answer: B

Q999. What is the purpose of Apache Airflow in big data workflows?

A. It provides air conditioning for data centers
B. Airflow is a workflow orchestration platform that schedules, monitors, and manages complex data pipeline dependencies as directed acyclic graphs
C. It processes data in real-time
D. It replaces Apache Spark

Answer: B

Q1000. What is the concept of data locality in distributed computing?

A. Storing all data in one location
B. Moving computation to where the data resides rather than moving large datasets to the compute nodes, minimizing network transfer
C. Data must always be moved to a central server
D. Data locality only matters for small datasets

Answer: B

Q1001. What is the difference between concept drift and data drift?

A. They are identical phenomena
B. Data drift is a change in input feature distributions while concept drift is a change in the relationship between features and the target variable
C. Concept drift only affects deep learning
D. Data drift does not require monitoring

Answer: B

Q1002. How does the shadow deployment pattern work for ML models?

A. Models run in the dark
B. The new model receives production traffic and generates predictions alongside the existing model, but only the old model's predictions are served while the new model's performance is monitored
C. It replaces the model immediately
D. Shadow models never see real data

Answer: B

Q1003. What is the purpose of a model registry in MLOps?

A. A government registration system
B. A centralized repository that stores, versions, tracks, and manages ML model artifacts, metadata, and lifecycle stages
C. A list of model architectures
D. A training data storage system

Answer: B

Q1004. What is infrastructure as code and why is it important for MLOps?

A. Writing code on physical infrastructure
B. Defining and managing infrastructure through code files, enabling reproducible, version-controlled, and automated provisioning of ML environments
C. It only applies to web development
D. IaC replaces all manual processes

Answer: B

Q1005. How does model compression benefit edge deployment?

A. It makes models larger
B. Techniques like pruning, quantization, and distillation reduce model size and computation, enabling deployment on resource-constrained edge devices like mobile phones and IoT sensors
C. Edge devices have unlimited resources
D. Compression always reduces accuracy unacceptably

Answer: B

Q1006. What is the role of feature flags in ML model deployment?

A. Flags used in feature engineering
B. Feature flags enable gradually enabling new models for specific user segments, allowing controlled rollout and quick disabling if issues arise
C. They only work for software features
D. Feature flags are identical to A/B tests

Answer: B

Q1007. What is the difference between online and offline model evaluation?

A. They produce identical results
B. Offline evaluation uses held-out historical data while online evaluation measures performance on live production traffic with real user interactions
C. Online evaluation is always more accurate
D. Offline evaluation uses production data

Answer: B

Q1008. What is the purpose of model explainability tools in production?

A. To make models more complex
B. To provide human-understandable explanations for individual predictions, supporting debugging, regulatory compliance, and user trust
C. Explainability is only for research
D. All models are inherently explainable

Answer: B

Q1009. What is the concept of model staleness in production?

A. Models that smell bad
B. A model becoming less accurate over time as the real-world data distribution changes from the data it was trained on
C. Models always improve with age
D. Staleness only affects neural networks

Answer: B

Q1010. What is the concept of algorithmic recourse and why is it important?

A. Recourse means repeating algorithms
B. Algorithmic recourse provides individuals with actionable steps to change an unfavorable AI decision, ensuring people are not permanently harmed by automated systems
C. It is a type of optimization algorithm
D. Recourse is only relevant for recommendation systems

Answer: B

Q1011. How can feedback loops amplify bias in AI systems?

A. Feedback loops always reduce bias
B. Biased AI decisions influence future data collection which reinforces the original bias, creating a cycle where the system becomes increasingly biased over time
C. Feedback loops do not exist in AI
D. They only affect recommendation systems

Answer: B

Q1012. What is the difference between explainability and interpretability in AI?

A. They are identical concepts
B. Interpretability means a model is inherently understandable while explainability provides post-hoc explanations for complex models that are not inherently transparent
C. Explainability is always preferred
D. Interpretability only applies to neural networks

Answer: B

Q1013. What is the concept of privacy-preserving machine learning?

A. Making ML models private property
B. Techniques that enable training and inference on sensitive data while protecting individual privacy, including federated learning, differential privacy, and secure computation
C. Privacy cannot be preserved in ML
D. It only applies to healthcare data

Answer: B

Q1014. What are model cards and why are they important?

A. Physical cards with model images
B. Standardized documentation describing a model's intended use, performance metrics, limitations, and ethical considerations, promoting transparency and responsible deployment
C. They are used for trading models
D. Model cards replace all other documentation

Answer: B

Q1015. What is the problem of proxy discrimination in AI systems?

A. Using proxy servers for AI
B. When AI uses seemingly neutral features that are highly correlated with protected attributes like race or gender, resulting in discriminatory outcomes despite not explicitly using those attributes
C. Proxy discrimination cannot occur in AI
D. It only happens with simple models

Answer: B

Q1016. How does the concept of AI safety differ from AI security?

A. They are identical concerns
B. AI safety focuses on preventing unintended harmful behaviors from AI systems themselves, while AI security protects AI systems from external malicious attacks
C. Safety is only about physical robots
D. Security is not relevant to AI

Answer: B

Q1017. What is the right to erasure and how does it apply to AI systems?

A. Erasing AI code permanently
B. Under GDPR, individuals can request deletion of their personal data, which creates challenges for AI systems trained on that data as the model may retain learned patterns
C. It does not apply to AI
D. Only applies to social media

Answer: B

Q1018. What is the environmental impact of training large AI models?

A. AI training has zero environmental impact
B. Training large models consumes significant energy and produces carbon emissions, with some large model training runs emitting as much CO2 as multiple transatlantic flights
C. Environmental impact is negligible
D. Only GPU manufacturing has environmental impact

Answer: B

Q1019. What is the concept of AI value alignment?

A. Aligning data values in a database
B. Ensuring AI systems' goals, behaviors, and decisions are consistent with human values, intentions, and ethical principles
C. Aligning model parameters to specific values
D. Value alignment is not important for AI

Answer: B

Q1020. What is the purpose of a staging environment in ML deployment?

A. A theater stage for AI presentations
B. A pre-production environment that mirrors production for testing models with realistic data and infrastructure before live deployment
C. A storage area for old models
D. A training environment for data scientists

Answer: B

Hard Questions

510 questions

Q1021. The Chinese Room argument by John Searle challenges:

A. Standard statistical data analytics methods
B. Strong AI - that machines can truly understand
C. Modern machine learning training algorithms
D. Weak AI - that machines can mimic intelligence

Answer: B

Q1022. Which of the following problems is considered AI-complete?

A. Sorting a list of numbers
B. Computing matrix multiplication
C. Performing a binary search
D. Natural language understanding

Answer: D

Q1023. The frame problem in AI refers to:

A. Challenges of reducing network latency across distributed systems
B. Complications of managing memory allocation in modern computers
C. Difficulty in determining what changes and what stays the same after an action
D. Problems with optimizing database indexing for query performance

Answer: C

Q1024. Which approach to AI attempts to mimic biological neural structures?

A. Symbolism
B. Evolutionary computation
C. Bayesian methods
D. Connectionism

Answer: D

Q1025. In the context of AI, what is the combinatorial explosion?

A. Rapid growth of possible solutions making brute force infeasible
B. An advanced technique for compressing large data files
C. A sudden failure in hardware caused by electrical overload
D. A specialized type of deep neural network architecture

Answer: A

Q1026. What distinguishes Artificial General Intelligence (AGI) from Narrow AI?

A. AGI consumes significantly less working memory
B. AGI requires absolutely no input training data
C. AGI can perform any intellectual task a human can
D. AGI is faster at one specific computational task

Answer: C

Q1027. The symbol grounding problem in AI concerns:

A. How to optimize the speed of sorting algorithms
B. How to efficiently compress large data archives
C. How to increase the clock speed of a processor
D. How symbols in an AI system get their meaning

Answer: D

Q1028. Which of the following is a characteristic of a multi-agent system?

A. Only one autonomous decision-maker
B. Multiple interacting intelligent agents
C. A single centralized processing unit
D. No communication between components

Answer: B

Q1029. What is the knowledge representation bottleneck in AI?

A. Slow data transfer speeds across network connections
B. Difficulty in encoding real-world knowledge into a formal system
C. Insufficient physical storage space on server hardware
D. Limited pixel resolution on visual display monitors

Answer: B

Q1030. In the context of data analytics, what is the CRISP-DM model?

A. A standard process model for data mining projects
B. A multi-layer neural network architecture design
C. A general-purpose interpreted programming language
D. A scalable relational database management system

Answer: A

Q1031. The Hessian matrix contains:

A. Eigenvalues of the matrix only
B. The determinant values only
C. Second-order partial derivatives
D. First-order derivatives only

Answer: C

Q1032. The Kullback-Leibler divergence measures:

A. The spread or variance of a single distribution
B. The central tendency or mean of a distribution
C. How one probability distribution differs from another
D. The linear correlation between two variables

Answer: C

Q1033. A Jacobian matrix represents:

A. First-order partial derivatives of a vector-valued function
B. Only the output values of a single scalar function
C. Only the second-order derivatives of a scalar function
D. Only the eigenvalues of a square transformation matrix

Answer: A

Q1034. In the context of ML, why is the log-likelihood used instead of likelihood?

A. It requires significantly less memory during training
B. It converts products to sums, making computation easier
C. It produces more statistically accurate final estimates
D. It runs faster on modern parallel computing hardware

Answer: B

Q1035. The singular value decomposition (SVD) decomposes a matrix into:

A. U, Sigma, and V-transpose matrices
B. Only the eigenvector components
C. Exactly two factor matrices
D. Only the eigenvalue components

Answer: A

Q1036. A saddle point in optimization is where:

A. The objective function reaches its overall global minimum
B. The gradient of the function is completely undefined
C. The objective function reaches its overall global maximum
D. The gradient is zero but it is neither a minimum nor maximum

Answer: D

Q1037. The moment generating function uniquely determines:

A. Only the median statistic
B. Only the expected mean value
C. The probability distribution
D. Only the variance estimate

Answer: C

Q1038. In matrix calculus, the gradient of x^T A x with respect to x is:

A. 2x
B. Ax
C. A^T
D. (A + A^T)x

Answer: D

Q1039. The condition number of a matrix indicates:

A. The trace or diagonal sum of the matrix
B. Sensitivity of the solution to small changes in input
C. The total number of columns in the matrix
D. The total number of rows in the matrix

Answer: B

Q1040. Jensen's inequality states that for a convex function f:

A. f(E[X]) "d E[f(X)]
B. f(E[X]) "e E[f(X)]
C. f(E[X]) = 0
D. f(E[X]) = E[f(X)]

Answer: A

Q1041. What is the Global Interpreter Lock (GIL) in Python?

A. A file system locking mechanism for preventing concurrent write access
B. A garbage collection tool for automated memory resource management
C. A cryptographic security feature for encrypting sensitive runtime data
D. A mutex preventing multiple threads from executing Python bytecode simultaneously

Answer: D

Q1042. Which tool is best for parallel processing of large datasets in Python?

A. Flask
B. Django
C. Tkinter
D. Dask

Answer: D

Q1043. What is the difference between deepcopy and copy in Python?

A. deepcopy creates copies of nested objects recursively, copy does not
B. copy is faster and creates a more thorough deep duplication
C. deepcopy only works on list objects and not dictionaries
D. They are completely identical in behavior and performance

Answer: A

Q1044. In NumPy, what does np.einsum() do?

A. Performs only element-wise operations on flat one-D arrays
B. Performs Einstein summation for multi-dimensional array operations
C. Creates identity matrices of a specified dimension size
D. Calculates eigenvalues of a square matrix decomposition

Answer: B

Q1045. What is vectorization in the context of NumPy?

A. Converting natural language text into numeric vectors
B. Performing operations on entire arrays instead of loops
C. Creating scalable vector graphics for visualization
D. A lossless data compression encoding technique

Answer: B

Q1046. What does the __slots__ attribute do in Python classes?

A. Defines abstract methods for interface contracts
B. Enables multiple inheritance across class chains
C. Restricts instance attributes and reduces memory usage
D. Creates new class-level methods from descriptors

Answer: C

Q1047. Which library provides GPU-accelerated computing for Python ML?

A. CuPy
B. Pandas
C. Matplotlib
D. SciPy

Answer: A

Q1048. What is a context manager in Python and why is it useful in ML?

A. It creates persistent database connections for data loading
B. It manages resources using with statements ensuring proper cleanup
C. It handles asynchronous network requests for API access
D. It manages the training context of machine learning models

Answer: B

Q1049. What is the purpose of __call__ in a Python class?

A. Copies the object into a new reference
B. Initializes a new class from a template
C. Destroys the instance and frees memory
D. Makes an instance callable like a function

Answer: D

Q1050. How does memory mapping (np.memmap) help in handling large datasets?

A. It maps files to memory allowing access without loading the entire file
B. It encrypts data at rest to ensure privacy and compliance
C. It compresses data into smaller archive files for disk savings
D. It duplicates data across servers for redundancy and backups

Answer: A

Q1051. What is the SMOTE technique used for?

A. Reducing the dimensionality of the feature space with PCA
B. Oversampling the minority class by creating synthetic examples
C. Undersampling the majority class by randomly removing records
D. Selecting only the most informative features from the dataset

Answer: B

Q1052. When should you use target encoding instead of one-hot encoding?

A. When the overall dataset is small
B. When there are only two categories present
C. When categorical variables have high cardinality
D. When the data is purely numerical

Answer: B

Q1053. What is data leakage in preprocessing?

A. When data columns are encrypted for secure storage
B. When data records are accidentally duplicated twice
C. When data is physically lost during the cleaning step
D. When information from the test set influences training

Answer: D

Q1054. What is the purpose of Yeo-Johnson transformation?

A. To detect and remove extreme outlier values from the numerical columns
B. To encode high-cardinality categorical features into integer representations
C. To make data more normally distributed, handling both positive and negative values
D. To fill in missing values using the mean or median of each column

Answer: C

Q1055. In handling missing data, what is MCAR?

A. Missing data that follows a clear and observable systematic pattern
B. Missing Completely At Random - missingness is unrelated to any variable
C. Missing data that occurs in only one specific column of data
D. Missing data that was caused by systematic data entry errors

Answer: B

Q1056. What is the KNN imputation method?

A. Replacing missing values with randomly generated numerical estimates
B. Deleting every row that contains any missing values from the dataset
C. Filling all missing entries with the single overall global mean value
D. Using K nearest neighbors to estimate missing values based on similar records

Answer: D

Q1057. Why should standardization be fit only on training data?

A. Because training data is inherently more important
B. Because the computation is significantly faster
C. To prevent data leakage from test set statistics
D. Because the test data is a much smaller sample

Answer: C

Q1058. What is the Isolation Forest algorithm used for in preprocessing?

A. Automated feature selection
B. Anomaly and outlier detection
C. Data range normalization
D. Missing value imputation

Answer: B

Q1059. What is ordinal encoding and when is it preferred over one-hot encoding?

A. It assigns ordered integers to categories with natural ordering
B. It is exactly the same approach as basic label encoding
C. It removes categorical variables from the feature set entirely
D. It always creates separate binary columns for each category

Answer: A

Q1060. What is the effect of multicollinearity on preprocessing?

A. It has absolutely no effect on preprocessing or model training
B. Highly correlated features can cause instability in model coefficients
C. It always improves the overall accuracy of the trained model
D. It consistently reduces the total required model training time

Answer: B

Q1061. Simpson's paradox in EDA refers to:

A. A trend that appears in groups but reverses when groups are combined
B. A visualization error caused by incorrect axis scale calibration
C. A specific type of missing data pattern observed in large datasets
D. A random sampling method for selecting representative subsets

Answer: A

Q1062. What is the purpose of the Kolmogorov-Smirnov test?

A. To calculate the arithmetic mean of a data sample
B. To test if a sample comes from a specific distribution
C. To encode categorical data into numerical integers
D. To remove extreme outlier values from the dataset

Answer: B

Q1063. What is the curse of dimensionality in EDA?

A. There are too few features for the model to learn
B. There is simply too much data volume to process
C. Data becomes increasingly sparse as dimensions increase
D. The data is already too clean to analyze further

Answer: C

Q1064. Cramér's V statistic measures:

A. The skewness of a single distribution
B. Correlation between two numerical variables
C. The arithmetic mean of a given dataset
D. Association between two categorical variables

Answer: B

Q1065. What is the difference between correlation and causation in EDA?

A. Correlation measures association; causation means one variable directly affects another
B. Causation is a weaker relationship than correlation
C. Correlation always directly implies a causal link
D. They are completely identical concepts in statistics

Answer: B

Q1066. What is the Shapiro-Wilk test used for?

A. Testing for the presence of outlier values
B. Testing if a dataset is normally distributed
C. Testing for the existence of duplicate rows
D. Testing for the count of missing null values

Answer: B

Q1067. In EDA, what is a bimodal distribution?

A. A completely uniform distribution
B. A distribution with two distinct peaks
C. A distribution with no peaks at all
D. A distribution with one single peak

Answer: B

Q1068. What does the Durbin-Watson statistic test for?

A. Normality of the data
B. Homoscedasticity
C. Multicollinearity issues
D. Autocorrelation in residuals

Answer: B

Q1069. What is heteroscedasticity and why is it problematic?

A. Non-constant variance of residuals, violating regression assumptions
B. A specific type of variable correlation
C. A pattern of missing data values
D. Perfectly constant variance of all residuals

Answer: C

Q1070. What is the purpose of the Andrews curves visualization?

A. To display proportional breakdowns in a circular pie format
B. To visualize multivariate data as curves, each representing one observation
C. To create standard grouped and stacked bar charts for categories
D. To plot time series trend lines with seasonal decomposition

Answer: B

Q1071. What is the kernel trick in SVM?

A. Mapping data to a higher-dimensional space without explicit computation
B. A dimensionality reduction visualization projection technique
C. A wrapper-based method for iterative feature selection
D. A preprocessing data cleaning and normalization technique

Answer: A

Q1072. What is the VC dimension?

A. The total count of input features in the dataset
B. The step size or magnitude of the learning rate
C. A measure of the capacity and complexity of a classification model
D. The total number of training samples available

Answer: C

Q1073. In gradient boosting, what does each subsequent tree learn?

A. The residuals (errors) of the previous ensemble
B. Random patterns found within the noise
C. The original target values from the training data
D. Nothing new beyond the original input features

Answer: D

Q1074. What is the difference between hard and soft margin SVM?

A. Hard margin and soft margin are completely identical approaches
B. Hard margin is always the preferred choice over soft margin
C. Soft margin SVM never works well on real-world datasets
D. Hard margin allows no misclassification; soft margin allows some with a penalty

Answer: C

Q1075. What is the Representer Theorem?

A. A foundational theorem about efficient compressed data representation and storage formats
B. The optimal solution in kernel methods can be expressed as a linear combination of kernel evaluations at training points
C. A graphical visualization principle for plotting high-dimensional feature space relationships
D. A data warehousing method for efficiently storing columnar records in structured databases

Answer: B

Q1076. What is Platt scaling?

A. A method to calibrate SVM outputs into probabilities using a sigmoid function
B. A standardization feature scaling method using z-score transforms of values
C. A model combination ensemble technique using weighted averaging of outputs
D. A min-max data normalization technique for scaling feature ranges to a bound

Answer: A

Q1077. What is the effect of increasing the C parameter in SVM?

A. The model becomes much simpler and heavily regularized overall
B. The decision boundary margin becomes significantly wider
C. The model becomes more sensitive to individual points, risking overfitting
D. The amount of L2 regularization penalty increases significantly

Answer: C

Q1078. What is the difference between Gini impurity and entropy in decision trees?

A. Gini impurity is always the better choice for all tree problems
B. Both measure node impurity but Gini is computationally simpler; entropy uses logarithms
C. Entropy is always the better choice for all tree problems
D. They produce completely different tree structures every single time

Answer: B

Q1079. What is the Structural Risk Minimization principle?

A. Always using the simplest available model regardless of performance
B. Maximizing model complexity to fit all training data points exactly
C. Balancing empirical risk and model complexity to minimize generalization error
D. Minimizing only the training error and ignoring test performance

Answer: B

Q1080. In logistic regression, what does multicollinearity cause?

A. More accurate and well-calibrated probability estimates
B. Significantly faster convergence during the training process
C. Unstable and unreliable coefficient estimates with large standard errors
D. Consistently better and more accurate model predictions overall

Answer: C

Q1081. What is the bias-variance decomposition of ensemble methods?

A. Bagging primarily reduces variance; boosting primarily reduces bias
B. Neither technique affects the bias or variance components
C. Both techniques only reduce the variance component of error
D. Both techniques only reduce the bias component of error

Answer: A

Q1082. How does XGBoost handle regularization differently from traditional Gradient Boosting?

A. XGBoost uses no regularization at all in its training objective
B. Traditional Gradient Boosting has stronger regularization overall
C. They handle regularization in the exact same identical manner
D. XGBoost includes L1 and L2 regularization terms in its objective function

Answer: D

Q1083. What is the difference between XGBoost and LightGBM's tree growing strategy?

A. Neither of them actually uses tree models
B. XGBoost grows level-wise; LightGBM grows leaf-wise
C. They grow trees in an identical manner always
D. XGBoost uses leaf-wise; LightGBM uses level-wise

Answer: D

Q1084. In CatBoost, how are categorical features handled?

A. They are silently ignored during the training and inference steps
B. They are completely removed from the feature set before training
C. They must be manually one-hot encoded before model training begins
D. Using ordered target statistics with random permutations to avoid target leakage

Answer: D

Q1085. What is the effect of increasing the number of trees in a Random Forest?

A. It always causes the model to underfit the training data
B. Performance improves then plateaus; it generally does not overfit
C. It always causes the model to overfit the training data
D. Performance always decreases with each additional tree

Answer: B

Q1086. What is the role of the subsample parameter in Gradient Boosting?

A. It sets the step size or learning rate of the optimization
B. It selects which features are considered at each individual tree split
C. It introduces stochastic gradient boosting by using a fraction of samples per tree
D. It controls the maximum allowed depth of each individual tree

Answer: C

Q1087. How does the isolation mechanism work in Isolation Forest for ensemble anomaly detection?

A. Anomalies require significantly more random splits to isolate
B. The splits are entirely random and have no diagnostic meaning
C. Anomalies are isolated in fewer splits because they are rare and different
D. All data points require an exactly equal number of splits

Answer: C

Q1088. What is Bayesian Model Averaging?

A. Random model selection from the available ensemble
B. Simple majority voting across all base model predictions
C. Weighting ensemble models by their posterior probabilities
D. Using only the single highest-accuracy base model

Answer: C

Q1089. What is negative correlation learning in ensemble methods?

A. Using the exact same training data for every individual model
B. Removing highly correlated input features from the dataset
C. Training all base models to produce completely identical outputs
D. Encouraging base learners to make diverse errors through a penalty term

Answer: D

Q1090. How does DART (Dropouts meet Multiple Additive Regression Trees) improve boosting?

A. By using significantly deeper trees for each individual boosting iteration
B. By removing the least important features entirely from the training data
C. By randomly dropping trees during boosting iterations to prevent over-specialization
D. By adding many more trees to the overall boosted ensemble of the model

Answer: B

Q1091. What is the Gaussian Mixture Model (GMM)?

A. A tree-based decision boundary model for supervised classification prediction
B. A probabilistic model assuming data comes from a mixture of Gaussian distributions
C. A linear regression method for predicting continuous numerical target values
D. A filter-based feature selection technique using statistical hypothesis tests

Answer: B

Q1092. How does spectral clustering work?

A. Uses deep neural networks to learn cluster assignment mappings
B. Uses decision tree ensembles to group data into discrete clusters
C. Uses eigenvalues of a similarity matrix to reduce dimensionality before clustering
D. Applies K-Means clustering directly without any preprocessing transformation

Answer: C

Q1093. What is the UMAP algorithm?

A. A dimensionality reduction technique based on manifold learning and topology
B. A gradient-based numerical regression training technique
C. A supervised multi-class classification prediction method
D. An unsupervised density-based spatial clustering algorithm

Answer: A

Q1094. In DBSCAN, what are core points, border points, and noise points?

A. Noise points are the ones that form the densest most cohesive clusters
B. Core points have enough neighbors; border points are near core points; noise points are isolated
C. Core points are the extreme outliers that are furthest from all clusters
D. All three point types are treated exactly the same by the clustering algorithm

Answer: B

Q1095. What is the difference between hard and soft clustering?

A. Hard assigns each point to one cluster; soft assigns probabilities of belonging to each cluster
B. Soft clustering does not exist as a valid machine learning technique
C. Hard clustering is always the strictly better approach for all problems
D. They are completely identical approaches with no meaningful differences

Answer: A

Q1096. What is the Information Criterion (BIC/AIC) used for in clustering?

A. Performing feature selection for supervised model training
B. Measuring the internal purity of each individual data cluster
C. Calculating the pairwise distances between all data points
D. Selecting the optimal number of clusters by balancing fit and complexity

Answer: D

Q1097. What is Self-Organizing Map (SOM)?

A. An unsupervised neural network that produces a low-dimensional representation of data
B. A supervised classification model that predicts categorical labels
C. A gradient boosting algorithm that builds sequential tree models
D. A linear regression model that predicts continuous target values

Answer: A

Q1098. How does the OPTICS algorithm improve upon DBSCAN?

A. It requires explicitly specifying the number K of clusters
B. It only works with fully labeled and supervised training data
C. It is always computationally faster than every other algorithm
D. It handles clusters of varying density without requiring a fixed epsilon

Answer: D

Q1099. What is the manifold hypothesis in unsupervised learning?

A. All real-world data is always uniformly distributed in space
B. Data always naturally forms spherical well-separated clusters
C. All feature dimensions are always statistically independent
D. High-dimensional data often lies on a lower-dimensional manifold

Answer: D

Q1100. What is contrastive learning in the context of unsupervised representation learning?

A. A linear regression method for predicting numerical outputs
B. A purely supervised classification technique using labels
C. A density-based spatial clustering approach for grouping data
D. Learning representations by contrasting similar and dissimilar pairs

Answer: D

Q1101. What is the Matthews Correlation Coefficient (MCC)?

A. The exact same metric as standard classification accuracy for balanced data
B. A metric specifically designed for evaluating unsupervised clustering
C. A metric that is only applicable to multi-class classification problems
D. A balanced metric using all four confusion matrix values, ranging from -1 to 1

Answer: D

Q1102. What is the difference between Type I and Type II errors?

A. The two types are completely identical errors
B. Type I is a false positive; Type II is a false negative
C. Type I is a false negative; Type II is a false positive
D. Both are actually true positive outcomes

Answer: B

Q1103. What is nested cross-validation?

A. An outer CV loop for evaluation with an inner CV loop for hyperparameter tuning
B. A single random train-test split without any cross-validation folds
C. Cross-validation performed entirely without any held-out validation data
D. The same standard approach as regular single-loop K-fold cross-validation

Answer: A

Q1104. What is the Cohen's Kappa statistic?

A. A metric that accounts for agreement occurring by chance between predicted and actual
B. The exact same calculation as standard overall classification accuracy
C. A regression-only metric for evaluating continuous predictions
D. A feature importance measure for ranking input variables

Answer: A

Q1105. What is calibration in the context of model evaluation?

A. Whether predicted probabilities reflect true likelihoods of outcomes
B. The overall classification accuracy of the trained model
C. A data preprocessing step for cleaning raw data
D. The process of selecting the most relevant features

Answer: A

Q1106. What is the Brier score?

A. A feature importance measure for trees
B. The mean squared error of probabilistic predictions
C. A ranking metric for ordering search results
D. A clustering score for partition quality

Answer: B

Q1107. When should you use the PR-AUC instead of ROC-AUC?

A. For evaluating unsupervised clustering partition quality
B. When dealing with highly imbalanced datasets where positive class is rare
C. When the class distribution is perfectly balanced and equal
D. For evaluating regression models with continuous outputs

Answer: B

Q1108. What is the purpose of bootstrapping in model evaluation?

A. Performing automated feature selection and dimensionality reduction
B. Estimating confidence intervals for metrics by resampling with replacement
C. Running hyperparameter tuning using grid or random search
D. Increasing the total amount of available training data samples

Answer: B

Q1109. What is the difference between leave-one-out CV and K-fold CV?

A. They are exactly identical approaches with no meaningful differences whatsoever
B. LOOCV is always significantly faster computationally than K-fold approaches
C. K-fold always produces strictly more accurate results than LOOCV methods
D. LOOCV uses N-1 samples for training and 1 for testing each time; K-fold uses larger folds

Answer: D

Q1110. What is the expected calibration error (ECE)?

A. A regression metric measuring the average absolute difference in predictions
B. The exact same calculation as the standard log loss or cross-entropy loss metric
C. The weighted average of difference between predicted confidence and actual accuracy across bins
D. A clustering metric measuring the silhouette score across all partitions

Answer: C

Q1111. What is the Boruta algorithm for feature selection?

A. A multi-layer feedforward deep neural network architecture for prediction
B. A gradient-based linear regression method for continuous value estimation
C. A wrapper method using shadow features and Random Forest to find relevant features
D. A density-based unsupervised clustering algorithm for grouping similar data

Answer: C

Q1112. What is the difference between wrapper, filter, and embedded feature selection?

A. Wrappers evaluate subsets with a model; filters use stats; embedded methods learn during training
B. Filters use trained models to evaluate each candidate feature subset iteratively
C. Wrappers are always strictly the best approach for every feature selection problem
D. They are all identical approaches that produce the exact same feature subset selections

Answer: A

Q1113. What is the curse of dimensionality's impact on feature engineering?

A. More features require exponentially more data to avoid sparsity
B. The number of dimensions has no effect on model training
C. Fewer features always cause the model to severely underfit
D. More features always improve the model performance metrics

Answer: A

Q1114. What is mutual information in feature selection?

A. A measure of the dependency between two variables based on information theory
B. The linear correlation coefficient between two variables
C. The overall variance of a single variable in a dataset
D. The arithmetic mean of a single variable in a dataset

Answer: A

Q1115. What is feature crossing in deep learning?

A. Combining two or more features to create a new feature capturing their interaction
B. Normalizing features to have zero mean and unit variance
C. Removing individual features from the dataset one at a time
D. Visualizing features using scatter plots and histograms

Answer: A

Q1116. How does LASSO perform feature selection?

A. By using only simple correlation analysis
B. By removing features entirely at random
C. By ranking all features only by variance
D. By driving some feature coefficients exactly to zero through L1 regularization

Answer: C

Q1117. What is the difference between forward and backward feature selection?

A. Forward starts empty and adds features; backward starts with all and removes features
B. Forward selection actually removes features one at a time from model
C. Backward selection actually adds features one at a time to the model
D. Forward and backward selection are completely identical methods

Answer: D

Q1118. What is permutation importance?

A. Simply ranking features by their alphabetical column order
B. A method for sorting features by their variance values
C. A data augmentation technique for expanding training data
D. Measuring importance by shuffling each feature and observing the drop in performance

Answer: D

Q1119. What is the concept of information leakage in feature engineering?

A. Features that are redundant copies of other existing features
B. Features that have missing values in some of their data records
C. Features containing target info that would not be available at prediction time
D. Features that contain a large amount of measurement noise

Answer: C

Q1120. What is automated feature engineering (e.g., Featuretools)?

A. Generating features completely at random without any data context
B. Deleting features from the dataset to reduce dimensionality
C. Relying exclusively on manual hand-crafted feature creation processes
D. Using algorithms to auto-generate features from relational data via deep feature synthesis

Answer: D

Q1121. What is the dying ReLU problem?

A. ReLU causes gradients to explode during backpropagation
B. Neurons permanently output zero because they get stuck in the negative region
C. ReLU increases peak memory usage beyond available capacity
D. ReLU activation is computationally too slow for large networks

Answer: B

Q1122. What is the Lottery Ticket Hypothesis?

A. All neural network architectures are equally effective regardless of size
B. Sparse network architectures never work well for any real task
C. Dense networks contain sparse subnetworks that can match full network performance
D. Larger networks are always strictly better than smaller alternatives

Answer: C

Q1123. What is knowledge distillation?

A. Removing irrelevant features from the input feature space
B. Training a smaller student model to mimic a larger teacher model
C. Transferring database records between different storage systems
D. Augmenting the training data with synthetic generated samples

Answer: B

Q1124. What is the purpose of skip connections (residual connections)?

A. Removing entire layers from the network to reduce complexity
B. Skipping the data preprocessing step to speed up training
C. Reducing the size of the training dataset to save storage
D. Allowing gradients to flow directly by adding input to output of a block

Answer: D

Q1125. What is gradient clipping?

A. Removing all gradients entirely from the backpropagation computation
B. Setting all gradient values to exactly zero after each update step
C. Increasing the learning rate to speed up convergence time
D. Limiting gradient values to a maximum threshold to prevent exploding gradients

Answer: D

Q1126. What is the difference between SGD with momentum and Adam?

A. SGD momentum uses a fixed learning rate with velocity; Adam adapts per parameter
B. Adam optimizer never converges to a good solution on real data
C. SGD with momentum is always faster than Adam for every problem
D. They are completely identical optimization algorithms with no differences

Answer: A

Q1127. What is neural architecture search (NAS)?

A. Automated methods for finding optimal neural network architectures
B. Feature selection and dimensionality reduction algorithms
C. Manual hand-designed network architecture engineering process
D. Data preprocessing and cleaning pipeline automation tools

Answer: A

Q1128. What is the difference between pre-training and fine-tuning?

A. They are completely identical stages in the same training process
B. Fine-tuning always happens before the pre-training phase
C. Pre-training is designed specifically for very small datasets
D. Pre-training learns general representations; fine-tuning adapts to a specific task

Answer: D

Q1129. What is the effect of batch size on training?

A. Smaller batch sizes are always strictly worse for model generalization
B. Larger batch sizes are always strictly better for all training scenarios
C. Larger batches give smoother gradients but may generalize worse; smaller batches add helpful noise
D. Batch size has absolutely no measurable effect on training or generalization

Answer: C

Q1130. What is mixed-precision training?

A. Using both 16-bit and 32-bit floating point to speed up training while maintaining accuracy
B. Using only 8-bit integer arithmetic for all network computation steps
C. Using only 64-bit double-precision floats for maximum numerical precision
D. Training neural network models entirely without any GPU acceleration

Answer: A

Q1131. What is the difference between causal and bidirectional self-attention?

A. Causal attention only looks at previous positions; bidirectional looks at all positions
B. Bidirectional attention only looks backward at previous tokens in sequence
C. Causal and bidirectional attention are completely identical mechanisms
D. Causal attention actually looks at all positions in the input sequence

Answer: C

Q1132. What is the mode collapse problem in GANs?

A. The generator produces limited variety instead of the full data distribution
B. The model architecture has grown far too large for memory
C. The discriminator fails completely and cannot distinguish anything
D. The training process runs far too slowly to converge in time

Answer: A

Q1133. What is the Wasserstein GAN (WGAN) and how does it improve training?

A. It uses only convolutional layers without any dense layers
B. It uses the Wasserstein distance for more stable training gradients
C. It completely removes the discriminator from the architecture
D. It uses standard cross-entropy loss for the discriminator

Answer: B

Q1134. What is Neural ODE?

A. Modeling continuous-depth neural networks using ordinary differential equations
B. A generative adversarial network variant for image synthesis
C. A standard discrete-layer feedforward neural network architecture
D. An unsupervised density-based technique for data clustering

Answer: A

Q1135. What is the Flash Attention algorithm?

A. An IO-aware exact attention algorithm reducing memory from quadratic to linear
B. A data loading technique for prefetching training batches from disk
C. An approximate attention method that trades accuracy for modest speed
D. A new non-linear activation function for transformer hidden layers

Answer: A

Q1136. What is the difference between model parallelism and data parallelism?

A. Data parallelism splits the model layers across devices for training
B. Model parallelism splits the model across devices; data parallelism replicates model with split data
C. Model parallelism splits the data while keeping the full model on one device
D. They are completely identical distributed training strategies with no differences

Answer: B

Q1137. What is the concept of equivariance in CNNs?

A. Output is always identical regardless of any transformation of input
B. All layers in the network have exactly equal learned weight values
C. Output transforms in the same way as the input (e.g., shift in input causes shift in output)
D. The neural network architecture is perfectly symmetric across all layers

Answer: A

Q1138. What is a diffusion model?

A. A recurrent neural network architecture for sequence modeling
B. An unsupervised density-based clustering algorithm for data
C. A generative model that learns to reverse a gradual noising process
D. A generative adversarial network variant using two competing agents

Answer: C

Q1139. What is the Mixture of Experts (MoE) architecture?

A. A standard model ensemble using simple majority voting
B. A data preprocessing method for feature normalization
C. A single monolithic large network without any routing
D. A model using a gating network to route inputs to specialized sub-networks

Answer: D

Q1140. What is quantization in deep learning?

A. Increasing the overall model size by adding more trainable parameters
B. Applying data augmentation transforms to expand the training set
C. Reducing weight and activation precision to lower bit representations for efficiency
D. Adding more hidden layers to increase the depth of the network

Answer: C

Q1141. What is the difference between GPT and BERT architectures?

A. Both use exactly the same attention mechanism and training objective
B. GPT uses bidirectional attention; BERT uses causal left-to-right attention
C. They are completely identical transformer architectures with no differences
D. GPT uses causal attention for generation; BERT uses bidirectional attention for understanding

Answer: D

Q1142. What is byte-pair encoding (BPE) in tokenization?

A. A tokenization approach that only works at the word level
B. A tokenization strategy that works at sentence level
C. A tokenization method operating at character level only
D. A subword tokenization algorithm that iteratively merges the most frequent character pairs

Answer: B

Q1143. What is the perplexity metric in language models?

A. The total elapsed time required for training the model
B. The number of parameters or total size of the model
C. A measure of how well a model predicts a sample; lower is better
D. The total number of words in the vocabulary of the model

Answer: C

Q1144. What is retrieval-augmented generation (RAG)?

A. Combining retrieval with generation to ground outputs in external knowledge
B. An unsupervised clustering method for grouping similar documents
C. A purely generative approach relying entirely on internal knowledge
D. A purely retrieval approach returning documents without generation

Answer: A

Q1145. What is the problem of hallucination in large language models?

A. Generating plausible-sounding but factually incorrect or fabricated information
B. Using too much GPU memory during the inference computation
C. Producing completely empty outputs with no generated tokens
D. Generating text output at an excessively slow processing speed

Answer: A

Q1146. What is the concept of positional encoding in Transformers?

A. Normalizing the positional coordinates of data in feature space
B. Encoding the physical position of GPUs in a server rack cluster
C. Adding position info to embeddings since Transformers lack inherent sequence order
D. Sorting the input data samples by their index in the dataset

Answer: C

Q1147. What is cross-lingual transfer learning?

A. Translating all text first then training on the translated version
B. Ignoring language differences and treating all text the same
C. Training on one language and applying knowledge to another language
D. Training separate independent models for each individual language

Answer: C

Q1148. What is the difference between extractive and abstractive summarization?

A. Extractive generates completely new text from the source material
B. Abstractive only selects existing sentences from the original text
C. Extractive and abstractive summarization are identical methods
D. Extractive selects existing sentences; abstractive generates new sentences

Answer: C

Q1149. What is prompt engineering?

A. Cleaning and preprocessing the raw input training data
B. Modifying the internal neural network model architecture
C. Training the language model from scratch on new domain data
D. Designing effective input prompts to guide LLM behavior without fine-tuning

Answer: D

Q1150. What is the RLHF (Reinforcement Learning from Human Feedback) technique?

A. Unsupervised pre-training using masked language modeling objectives
B. Standard supervised fine-tuning using labeled input-output training pairs
C. Data augmentation using paraphrasing and back-translation methods
D. Fine-tuning language models using human preference judgments as reward signals

Answer: D

Q1151. What is the Feature Pyramid Network (FPN)?

A. A single-scale object detector that operates at only one fixed resolution level
B. A standard image classification network without multi-scale feature fusion
C. A generative adversarial network for producing synthetic realistic images
D. A multi-scale feature extractor combining low-res semantic and high-res spatial features

Answer: D

Q1152. What is the difference between one-stage and two-stage object detectors?

A. One-stage (YOLO) detects directly; two-stage (R-CNN) proposes regions then classifies
B. One-stage detectors are always significantly slower than two-stage methods
C. They are completely identical detection approaches with no performance differences
D. Two-stage detectors do not use region proposals at any processing stage

Answer: A

Q1153. What is panoptic segmentation?

A. A task that performs only semantic segmentation without instances
B. A task that performs only instance segmentation without semantics
C. A task that is limited to standard object detection only
D. A unified task combining semantic segmentation (stuff) and instance segmentation (things)

Answer: D

Q1154. What is the Vision Transformer (ViT)?

A. A recurrent neural network architecture designed for image sequences
B. A generative adversarial network for producing synthetic images
C. A specialized convolutional neural network variant for image processing
D. Applying the Transformer architecture directly to image patches for classification

Answer: D

Q1155. What is contrastive learning in computer vision (e.g., SimCLR)?

A. A two-stage region-proposal based object detection algorithm pipeline
B. Learning visual representations by contrasting augmented views of the same image against others
C. A fully supervised classification method that requires complete labeled training data
D. A generative adversarial network training method for synthesizing realistic images

Answer: B

Q1156. What is the concept of deformable convolutions?

A. Standard fixed-grid convolutions with no adaptive sampling offsets
B. Convolutions with learnable offsets that adapt the sampling grid to object shapes
C. One-by-one pointwise convolutions for changing channel depth
D. Dilated convolutions that increase the receptive field without offsets

Answer: B

Q1157. What is neural style transfer?

A. Classifying images into predefined categorical labels with CNNs
B. Applying the artistic style of one image to the content of another using CNNs
C. Training a brand new neural network from scratch on random data
D. Applying data augmentation transforms to expand the training set

Answer: B

Q1158. What is depth estimation in computer vision?

A. Predicting the distance of each pixel from the camera
B. Measuring the resolution of the image in pixels
C. Classifying the entire image into a category
D. Counting the total number of objects in a scene

Answer: A

Q1159. What is the CLIP model?

A. A model learning visual concepts from language supervision, connecting images and text
B. A single-modality object detector that only processes image pixel data
C. A standard supervised image classifier trained only on labeled image categories
D. A text-only language model without any visual understanding capabilities

Answer: A

Q1160. What is 3D point cloud processing?

A. Processing natural language text data from document sources
B. Processing unstructured 3D spatial data points from sensors like LiDAR
C. Processing continuous audio waveform signals from microphones
D. Processing standard two-dimensional image pixel data from cameras

Answer: B

Q1161. What is the Lambda architecture?

A. A single-server architecture running everything on one machine
B. An architecture supporting only offline batch processing workloads
C. An architecture supporting only real-time stream processing
D. A big data architecture combining batch and real-time stream processing

Answer: D

Q1162. What is Apache Flink?

A. A standalone relational database for transactional query workloads
B. A stream processing framework with event-time processing and exactly-once semantics
C. A batch-only data processor without any stream processing capability
D. An interactive data visualization and dashboard charting tool

Answer: B

Q1163. What is the difference between horizontal and vertical scaling?

A. Vertical scaling adds additional machines to the cluster pool
B. They are completely identical approaches to scaling infrastructure
C. Horizontal scaling upgrades individual machine hardware resources
D. Horizontal adds more machines; vertical adds more power to existing machines

Answer: D

Q1164. What is the concept of data lineage?

A. The specific file format of the data such as CSV or JSON
B. Tracking data from its origin through all transformations to its current state
C. The chronological age of the data since it was created
D. The total size of the data measured in bytes or rows

Answer: B

Q1165. What is eventual consistency in distributed systems?

A. All replicas converge to the same value given enough time without new updates
B. Consistency is fundamentally impossible in any distributed system
C. All replicas are immediately and perfectly consistent at all times
D. Data is never consistent across any of the distributed replicas

Answer: A

Q1166. What is Apache Arrow?

A. A cross-language platform for in-memory columnar data with zero-copy reads
B. A distributed event streaming platform for real-time pipelines
C. A supervised machine learning library for model training
D. A standalone relational database for transactional query workloads

Answer: A

Q1167. What is the Kappa architecture?

A. A simplified Lambda alternative using only stream processing for real-time and batch
B. An architecture designed only for small-scale single-machine data
C. The exact same architecture as Lambda with no simplification
D. An architecture supporting only offline batch processing workloads

Answer: A

Q1168. What is data skew in distributed processing?

A. Perfectly even and balanced distribution of data across all nodes
B. Uneven data distribution across partitions causing some nodes to be overloaded
C. Data records that are missing or have null values throughout
D. Data records that have been accidentally duplicated in storage

Answer: B

Q1169. What is the role of a distributed consensus algorithm like Raft or Paxos?

A. Visualizing distributed system metrics on dashboards
B. Sorting large distributed data records across nodes
C. Training machine learning models on distributed nodes
D. Ensuring all nodes agree on shared state despite failures

Answer: D

Q1170. What is columnar storage and why is it efficient for analytics?

A. A type of secondary database index for accelerating query lookups
B. Storing data by rows as in traditional relational database table layouts
C. Storing data by columns not rows, enabling efficient compression and selective reads
D. A data visualization technique for displaying column chart graphics

Answer: C

Q1171. What is the concept of shadow deployment?

A. Running a new model in parallel with production without serving its predictions
B. Deploying the model at night during off-peak hours only
C. A/B testing between two model versions with live traffic
D. Canary deployment with a small percentage of live users

Answer: A

Q1172. What is concept drift vs data drift?

A. Data drift only affects the target output variable distribution
B. Concept drift only affects the distribution of input features alone
C. Concept drift changes input-target relationship; data drift changes input distribution
D. They are completely identical phenomena with no meaningful differences

Answer: C

Q1173. What is model explainability in production?

A. Storing more raw training data for future retraining sessions
B. Providing interpretable explanations for why a model makes specific predictions
C. Making the model inference significantly faster and more efficient
D. Reducing the overall size and memory footprint of the model

Answer: B

Q1174. What is the purpose of experiment tracking tools like MLflow?

A. Recording parameters metrics code and artifacts for each experiment
B. Only visualizing final model results on a dashboard
C. Only storing the raw training data on a file system
D. Only deploying models to a production serving endpoint

Answer: A

Q1175. What is the blue-green deployment strategy?

A. A data visualization technique using a blue-green color palette
B. Running models on GPUs with blue and green LED indicator lights
C. A model training strategy using two alternating learning rates
D. Maintaining two identical production environments and switching traffic between them

Answer: D

Q1176. What is model compression for deployment?

A. Increasing numerical precision of weights from FP16 to FP64
B. Adding more hidden layers to increase the depth of network
C. Reducing model size and complexity while maintaining acceptable performance
D. Making models significantly larger by adding trainable parameters

Answer: C

Q1177. What is the role of infrastructure as code (IaC) in MLOps?

A. Collecting and gathering raw data from external data sources
B. Writing machine learning algorithm implementations and model code
C. Defining and managing ML infrastructure through code for reproducibility
D. Training machine learning models on labeled training datasets

Answer: C

Q1178. What is a serving graph in ML deployment?

A. A computation graph used only during the model training process
B. An interactive data visualization chart for exploring model outputs
C. A diagram showing the internal layers and neurons of a network
D. A DAG of preprocessing inference and postprocessing steps during prediction

Answer: D

Q1179. What is the challenge of training-serving skew?

A. The training dataset being too small for adequate model learning overall
B. The model training process converging far too slowly to be practical
C. Models becoming too large to fit within the available GPU memory capacity
D. Differences between training and serving environments causing inconsistent predictions

Answer: B

Q1180. What is GitOps for ML?

A. Using Git as single source of truth for ML infrastructure configs and deployments
B. Only storing the Python source code files in a Git repository
C. A gradient-based training method for optimizing neural networks
D. A type of version control system different from standard Git

Answer: A

Q1181. What is the alignment problem in AI?

A. Ensuring advanced AI goals and behaviors align with human values and intentions
B. Reducing the computational costs of training large-scale AI models
C. Making AI systems process data at significantly faster inference speeds
D. Training machine learning models to achieve the highest accuracy metrics

Answer: A

Q1182. What is membership inference attack?

A. Deleting training data from the model's storage permanently
B. Stealing the complete set of model weights from the server
C. Determining whether a specific data point was used in training a model
D. Changing the model's predictions by modifying its architecture

Answer: C

Q1183. What is the concept of Fairness through Awareness?

A. Completely ignoring all protected attributes during model prediction
B. Making random predictions without considering any input features
C. Using the exact same model configuration for every single individual
D. Treating similar individuals similarly using a task-specific similarity metric

Answer: D

Q1184. What is model inversion attack?

A. Improving the model's overall prediction accuracy
B. Reconstructing training data features from model outputs
C. Making the model inference significantly faster
D. Augmenting the training data with synthetic samples

Answer: B

Q1185. What is the difference between equality and equity in AI fairness?

A. Equity ignores all differences between individuals and groups
B. Equality is always strictly the better approach for fairness goals
C. They are completely identical concepts with no meaningful differences
D. Equality gives the same treatment; equity adjusts treatment for fair outcomes

Answer: D

Q1186. What is the EU AI Act?

A. A supervised machine learning algorithm for classification tasks
B. A structured data format for storing model configurations
C. A risk-based regulatory framework for AI in the European Union
D. A general-purpose compiled programming language specification

Answer: C

Q1187. What is algorithmic recourse?

A. Providing individuals with actionable steps to change an unfavorable AI decision
B. Completely removing the algorithm from production deployment
C. Using a completely different dataset for model retraining
D. Making the algorithm run faster through hardware optimization

Answer: A

Q1188. What is the difference between disparate treatment and disparate impact?

A. Disparate treatment is the unintentional adverse effect on specific groups
B. Disparate treatment uses protected attributes intentionally; disparate impact causes unequal effects
C. Disparate impact is the intentional use of protected attributes in a decision model
D. They are completely identical legal concepts with no meaningful differences

Answer: C

Q1189. What is homomorphic encryption and how does it relate to AI privacy?

A. A compression algorithm for reducing data file sizes
B. A process of permanently deleting data from storage
C. Encryption allowing computation on encrypted data without decrypting it
D. A standard symmetric encryption method for securing files

Answer: C

Q1190. What is the concept of AI safety?

A. Making AI systems process data at significantly faster speeds for better user experience
B. Ensuring AI systems behave as intended without causing unintended harm as capabilities grow
C. Marketing and promoting AI products and services to potential commercial customers
D. Reducing the overall financial costs of training and deploying large AI model systems

Answer: B

Q1191. A model achieves 99% training accuracy but only 55% test accuracy. Which strategy best addresses this?

A. Increase model complexity by adding many more parameters and layers
B. Remove all validation data to maximize available training data volume
C. Increase the learning rate to help the model converge much faster
D. Apply regularization techniques and collect more diverse training data

Answer: D

Q1192. When labeled data is scarce but unlabeled data is abundant, which learning paradigm is most suitable?

A. Rule-based expert system with manually crafted knowledge base rules
B. Pure unsupervised learning ignoring all available labeled data points
C. Semi-supervised learning combining labeled and unlabeled data sources
D. Fully supervised learning with extensive data augmentation pipelines

Answer: C

Q1193. Which challenge arises when deploying ML models trained on historical data that no longer reflects reality?

A. Gradient vanishing causing training to stall at suboptimal solutions
B. Feature explosion resulting from too many correlated input variables
C. Memory overflow due to the increasing size of the stored model files
D. Data drift causing degraded model performance over time in production

Answer: D

Q1194. A startup wants to predict rare diseases from patient records. What is the primary ML challenge?

A. Class imbalance where positive cases are extremely underrepresented
B. Feature extraction is impossible from electronic health record data
C. The data will be too structured for machine learning algorithms used
D. The model will require too many GPU hours for simple tabular data

Answer: A

Q1195. Why is the No Free Lunch theorem significant in ML algorithm selection?

A. It states no single algorithm performs best across every possible task
B. It proves deep learning always outperforms all traditional ML methods
C. It shows unsupervised learning is superior to supervised approaches
D. It guarantees ensemble methods will always improve baseline accuracy

Answer: A

Q1196. For autonomous vehicles, which combination of AI subfields is most critical?

A. Recommendation systems combined with collaborative filtering engines
B. Computer vision with reinforcement learning and sensor fusion systems
C. Sentiment analysis combined with topic modeling for understanding
D. Natural language processing combined with text summarization methods

Answer: B

Q1197. What is the key difference between model-based and instance-based learning?

A. Model-based is always faster at inference while instance-based is always faster
B. Model-based builds explicit parameter models while instance-based memorizes examples
C. Model-based requires no training while instance-based needs extensive training time
D. Model-based only works with images while instance-based only handles text data

Answer: B

Q1198. Adding more training data yields no accuracy improvement. What does this indicate about the model?

A. The computing hardware has reached its maximum processing speed limits
B. The model has reached its capacity and needs a more complex architecture
C. The learning rate is set too high causing the optimizer to diverge rapidly
D. The dataset contains no useful patterns for any learning algorithm to find

Answer: B

Q1199. In AI ethics, what does algorithmic accountability primarily require from organizations?

A. Explaining and taking responsibility for AI-driven decisions outcomes
B. Publishing all source code as open source for public use and changes
C. Restricting AI development to only government-approved research labs
D. Mandating that all AI models achieve above ninety-five percent rate

Answer: A

Q1200. Which approach best addresses catastrophic forgetting in continual learning systems?

A. Removing all previous training data before introducing new examples
B. Reducing the model size to prevent storing too much old information
C. Using elastic weight consolidation to protect important parameters
D. Retraining the entire model from scratch every time new data arrives

Answer: C

Q1201. Why is the Hessian matrix important in second-order optimization methods?

A. It captures second-order derivative information to estimate curvature
B. It stores the training data samples needed for batch processing
C. It determines the initial random weight values for network layers
D. It selects the appropriate activation function for each neuron

Answer: A

Q1202. When KL divergence between two distributions is zero, what does this indicate?

A. One distribution has zero variance while the other has infinite range
B. The two distributions have no overlapping support in their domains
C. Both distributions are perfectly negatively correlated with each other
D. The two probability distributions are completely identical everywhere

Answer: D

Q1203. Why can saddle points be problematic for optimization in high-dimensional spaces?

A. Saddle points make it impossible to compute any derivative at that point
B. Saddle points always cause the loss function value to increase unboundedly
C. Gradients near saddle points approach zero causing very slow convergence
D. Saddle points only occur in one-dimensional optimization never in higher

Answer: C

Q1204. What is the mathematical justification for using cross-entropy loss in classification?

A. It guarantees convergence in exactly one epoch of training every time
B. It derives from maximum likelihood estimation under a Bernoulli model
C. It eliminates the need for any regularization in the model training
D. It is computationally simpler than all other available loss functions used

Answer: B

Q1205. Why does the curse of dimensionality degrade distance-based algorithms?

A. Higher dimensions always require more memory than any hardware provides
B. Higher dimensions guarantee all data points will be perfectly separable
C. In high dimensions the class count always exceeds the sample count total
D. In high dimensions all pairwise distances converge making neighbors meaningless

Answer: D

Q1206. What role does the Jacobian matrix play in training multi-output neural networks?

A. It contains all first-order partial derivatives of outputs versus inputs
B. It determines the optimal batch size for stochastic gradient descent
C. It stores bias terms for each layer of the neural network model used
D. It specifies architecture including number of layers and their widths

Answer: A

Q1207. In Bayesian inference, how does the prior influence the posterior when data is very limited?

A. The prior has minimal influence and the posterior matches the likelihood
B. The prior is completely ignored by all Bayesian inference methods used
C. The prior dominates the posterior since limited data provides weak signal
D. The prior and posterior are always identical regardless of observed data

Answer: C

Q1208. What mathematical property makes softmax suitable for multi-class classification output?

A. It compresses all inputs to a fixed range of negative one to one
B. It outputs values that always sum to exactly one across all classes
C. It always produces integer values representing discrete class indices
D. It ensures every output is either exactly zero or exactly one always

Answer: B

Q1209. Why is the Fisher Information Matrix used in natural gradient descent?

A. It captures statistical manifold curvature for more efficient update steps
B. It eliminates the requirement for computing any first-order derivatives
C. It replaces the need for any loss function during model training completely
D. It guarantees global convergence regardless of the model architecture used

Answer: A

Q1210. What is the significance of positive semi-definiteness in covariance matrices?

A. It restricts the matrix to having exactly two distinct eigenvalue values
B. It ensures the matrix can only contain integer values in all its entries
C. It means the matrix inverse always exists and is unique for every case
D. It guarantees all eigenvalues are non-negative representing valid variances

Answer: D

Q1211. For a custom scikit-learn transformer, which methods must be implemented for pipeline compatibility?

A. Only the predict method is required since transformers always make predictions
B. The score and evaluate methods are the only ones required for all pipelines
C. Only the transform method needs to be implemented for pipeline compatibility
D. The fit and transform methods must both be implemented at minimum for use

Answer: D

Q1212. What is the advantage of multiprocessing over threading for CPU-bound ML tasks in Python?

A. Multiprocessing creates separate processes bypassing GIL for true parallelism
B. Threading bypasses the GIL while multiprocessing is limited by the GIL lock
C. Threading uses less memory because it creates separate isolated memory spaces
D. Multiprocessing is always slower but provides much better error handling now

Answer: A

Q1213. Why should you use np.allclose() instead of == when comparing floating-point arrays?

A. The allclose function is significantly faster than equality operator for arrays
B. The equality operator cannot be applied to any NumPy arrays at all ever
C. Floating-point arithmetic introduces rounding errors making exact comparison fail
D. The equality operator only works for integer arrays and raises errors on floats

Answer: C

Q1214. Why is it critical to apply fit() only on training data for preprocessing in production ML?

A. Fitting on test data would make preprocessing steps run much more slowly
B. Fitting on all data causes data leakage by incorporating test set statistics
C. Fitting on training data alone ensures the model always achieves higher loss
D. Fitting on test data would automatically delete the test set from memory

Answer: B

Q1215. What problem does functools.lru_cache solve in feature engineering pipelines?

A. It logs all function calls to a database for auditing and tracking uses
B. It converts all function outputs to lower precision to save system memory
C. It automatically distributes computation across multiple GPU device units
D. It caches expensive function results to avoid redundant recomputation work

Answer: D

Q1216. When handling very large datasets in pandas, what approach reduces memory most effectively?

A. Specifying optimal dtypes and reading data in chunks with iteration
B. Removing all column names to reduce the metadata overhead in RAM
C. Sorting the DataFrame by index before performing any analysis steps
D. Converting all columns to string type to ensure uniform representation

Answer: A

Q1217. What is the purpose of __getitem__ and __len__ in a PyTorch custom Dataset class?

A. They define how to save and load model checkpoints to and from disk
B. They specify loss function and optimizer used during backpropagation
C. They enable indexed data access and size reporting for the DataLoader
D. They control the learning rate schedule during model training loop runs

Answer: C

Q1218. Why is np.einsum() considered more flexible than standard NumPy operations for tensors?

A. It runs exclusively on GPU hardware while standard operations use CPU only
B. It expresses complex multi-dimensional operations in compact Einstein notation
C. It automatically converts all inputs to higher precision floating point types
D. It only works with two-dimensional arrays making the interface much simpler

Answer: B

Q1219. What design pattern does scikit-learn's Pipeline class implement and why is it beneficial?

A. The observer pattern notifying all components when training data changes
B. The composite pattern chaining steps to prevent data leakage in CV folds
C. The singleton pattern ensuring only one model instance exists in memory
D. The factory pattern creating new model instances based on input data type

Answer: B

Q1220. How does contextlib.contextmanager help manage resources in data processing?

A. It provides automatic type checking for all function arguments runtime
B. It creates permanent global variables accessible from any module at all
C. It enables resource management using generator-based context managers
D. It automatically converts synchronous code into asynchronous coroutines

Answer: C

Q1221. A dataset has 40% missing values in a critical feature. Which imputation strategy is most appropriate?

A. Ignore the feature entirely and remove it from the training dataset now
B. Simply delete all rows with missing values to ensure data quality overall
C. Replace all missing values with zero since it is the simplest approach
D. Use a predictive model like KNN or regression to impute the missing values

Answer: D

Q1222. Why can applying preprocessing transformations before splitting data cause data leakage?

A. Preprocessing transformations always modify the target variable directly
B. Statistics from test data influence the transformation applied to training data
C. Transformations always remove the most important features from the dataset
D. Preprocessing makes the model run slower during the inference stage overall

Answer: B

Q1223. When dealing with high-cardinality categorical features, why might target encoding outperform one-hot?

A. One-hot encoding creates massive sparse matrices while target encoding is compact
B. Target encoding always produces more accurate results regardless of the data
C. Target encoding removes the need for any validation set during training time
D. One-hot encoding automatically introduces data leakage into training process

Answer: A

Q1224. In time-series data, why must preprocessing respect temporal order during train-test splitting?

A. Temporal order only matters for visualization and does not affect model results
B. Time-series data must always be converted to images before model training steps
C. Random splitting can leak future information into training causing unrealistic scores
D. Time-series data always has fewer missing values than cross-sectional datasets

Answer: C

Q1225. A feature has extreme outliers that cannot be removed because they represent valid rare events. What approach is best?

A. Use mean imputation to replace all outlier values with the column average
B. Apply min-max scaling which will properly handle extreme outlier values well
C. Apply robust scaling using median and IQR which is resistant to outlier values
D. Delete the entire feature since it contains any outlier values in the data

Answer: C

Q1226. Why is the Yeo-Johnson transformation preferred over Box-Cox for some datasets?

A. Yeo-Johnson is always faster computationally than Box-Cox transformation
B. Yeo-Johnson only works with categorical features unlike Box-Cox method
C. Yeo-Johnson can handle both positive and negative values unlike Box-Cox
D. Yeo-Johnson produces exactly normal distributions while Box-Cox does not

Answer: C

Q1227. When preprocessing text data for NLP, why might subword tokenization outperform word-level tokenization?

A. Subword tokenization removes the need for any stopword removal preprocessing
B. It handles out-of-vocabulary words by breaking them into known subword units
C. It converts all text into numerical features without any vocabulary mapping
D. Subword tokenization always produces shorter sequences than word tokenization

Answer: B

Q1228. A dataset has features measured in different units with varying scales and distributions. Which preprocessing combination is most robust?

A. Apply no preprocessing since modern algorithms handle raw data effectively
B. Apply only min-max scaling to all features uniformly without further analysis
C. Use power transformation followed by standardization for each feature column
D. Use one-hot encoding on all features regardless of their data type values

Answer: C

Q1229. Why should you be cautious when using mean imputation for features with non-normal distributions?

A. It artificially reduces variance and can distort the true distribution shape
B. Mean imputation only works with categorical features not numerical ones
C. Mean imputation always increases the standard deviation of the feature values
D. It automatically removes all outliers before computing the replacement value

Answer: A

Q1230. In a pipeline processing streaming data, why is incremental preprocessing preferred over batch preprocessing?

A. It automatically handles all missing values without any configuration needed
B. It processes data point by point without needing the entire dataset in memory
C. Incremental preprocessing eliminates the need for any feature engineering step
D. Incremental preprocessing always produces more accurate transformations overall

Answer: B

Q1231. A dataset shows a Pearson correlation of 0.02 between two features, but the scatter plot reveals a clear U-shaped pattern. What explains this?

A. The data has too many missing values to compute correlation correctly
B. The scatter plot is incorrectly rendered due to a visualization error
C. A correlation of 0.02 actually indicates a very strong relationship here
D. Pearson only captures linear relationships and misses non-linear patterns

Answer: D

Q1232. When exploring a dataset with 500 features, which dimensionality reduction technique is most useful for initial EDA visualization?

A. Removing all features except the first two columns from the original data
B. Manually selecting 2 features based on column name alphabetical ordering
C. Using t-SNE or UMAP to project high-dimensional data into 2D for plotting
D. Creating 500 individual histograms and reviewing each one sequentially

Answer: C

Q1233. During EDA you discover that a feature has a bimodal distribution. What might this suggest about the underlying data?

A. The data may contain two distinct subpopulations mixed into one feature
B. Bimodal distributions always indicate that the feature is perfectly normal
C. The feature should be removed because bimodal distributions are always errors
D. The feature has too many outliers and needs to be replaced with its mean

Answer: A

Q1234. Why might the mean be a misleading measure of central tendency for income data?

A. Income data always follows a perfectly normal distribution in all nations
B. The mean cannot be computed for continuous variables like income amounts
C. Income data is always categorical so the mean has no mathematical meaning
D. Income data is typically right-skewed so the mean is pulled by high earners

Answer: D

Q1235. How does Simpson's Paradox affect conclusions drawn during exploratory data analysis?

A. It only occurs in datasets with fewer than one hundred total observations
B. A trend in subgroups can reverse when groups are combined into aggregate
C. It guarantees that adding more features will always improve model accuracy
D. It makes all statistical tests invalid regardless of sample size used here

Answer: B

Q1236. A feature shows near-zero variance in the training set. Why should this be investigated before modeling?

A. They carry almost no discriminative information and may cause instability
B. Near-zero variance features provide the strongest predictive signal always
C. Near-zero variance means the feature is perfectly normally distributed
D. They always indicate data collection errors that must be manually fixed

Answer: A

Q1237. You observe that two features have a Spearman correlation of 0.95. What preprocessing step might you consider?

A. Multiply the two features together to create a more powerful interaction term
B. Remove one of the redundant features or use PCA to reduce the collinearity
C. Convert both features to categorical variables to eliminate the correlation
D. Add both features to the model since high correlation always helps performance

Answer: B

Q1238. During EDA of a classification dataset, you notice heavy class overlap in feature space. What does this suggest?

A. The classification task is inherently difficult and may need complex models
B. The dataset is perfect for a simple linear classifier without any tuning
C. The features should all be removed and replaced with randomly generated ones
D. Class overlap always means the labels are incorrect and need relabeling

Answer: A

Q1239. Why is it important to analyze the relationship between features and the target variable separately for different subgroups?

A. Analyzing subgroups always produces the same results as aggregate analysis does
B. Different subgroups may exhibit different patterns that aggregate analysis hides
C. Subgroup analysis is never useful and only wastes computational resources
D. Subgroup analysis is only applicable to time-series data and no other types

Answer: B

Q1240. A dataset contains timestamps spanning five years. Which EDA technique best reveals seasonal patterns and trends?

A. Creating a single histogram of all values regardless of their timestamps
B. Sorting the data alphabetically by the timestamp column values only
C. Computing a single mean value across all five years of data points
D. Time-series decomposition into trend, seasonal, and residual components

Answer: D

Q1241. A linear SVM fails to classify a dataset with circular decision boundaries. Which approach best resolves this?

A. Switching to an RBF kernel to capture the non-linear circular boundary
B. Increasing the regularization parameter C to a very large positive value
C. Reducing the training set size to remove noise from the boundary area
D. Adding more linear features computed from the existing feature columns

Answer: A

Q1242. Why does Naive Bayes often perform surprisingly well despite its strong independence assumption?

A. The independence assumption is actually true for nearly all real-world datasets
B. Classification only needs correct ranking of class probabilities not exact values
C. Naive Bayes secretly uses feature correlations even though it assumes independence
D. The algorithm automatically corrects for violated assumptions during training

Answer: B

Q1243. In gradient boosting, how does each subsequent tree improve the overall model?

A. Each tree duplicates the first tree but with different random weight values
B. Each tree trains on entirely new data that was not used by any prior tree
C. Each tree is trained on a random subset of features ignoring previous results
D. Each tree fits the residual errors of the combined predictions of prior trees

Answer: D

Q1244. When should you prefer Ridge regression over Lasso regression for regularization?

A. When you need exactly half the features to have zero coefficient values only
B. When you want to perform automatic feature selection by zeroing coefficients
C. When most features are relevant and you want to shrink all coefficients evenly
D. When the dataset has categorical features that need one-hot encoding first

Answer: C

Q1245. A decision tree achieves perfect training accuracy but poor test accuracy. Which pruning strategy is most effective?

A. Grow the tree to maximum depth then add more features to improve test scores
B. Increase the minimum samples per leaf to the total training set size value
C. Remove all leaf nodes and keep only the root node for maximum simplicity
D. Apply cost-complexity pruning to find the subtree with best cross-validated score

Answer: D

Q1246. Why is the choice of distance metric critical for KNN performance on high-dimensional data?

A. In high dimensions most distance metrics lose discriminative power between points
B. High-dimensional data always has the same distances regardless of metric used
C. KNN only works with Euclidean distance so the choice is never actually needed
D. Distance metrics have no effect on KNN in any number of dimensions at all

Answer: A

Q1247. How does Platt scaling calibrate the output probabilities of an SVM classifier?

A. It fits a logistic regression on the SVM decision values to produce probabilities
B. It trains a separate neural network to convert SVM outputs into probabilities
C. It replaces the SVM kernel with a probability density estimation function
D. It applies min-max normalization directly to the raw SVM decision values

Answer: A

Q1248. In multi-class classification, what advantage does one-vs-rest have over one-vs-one strategy?

A. Both strategies always produce identical results regardless of the dataset
B. One-vs-rest trains fewer classifiers making it computationally more efficient
C. One-vs-one trains fewer classifiers making it computationally more efficient
D. One-vs-rest always achieves higher accuracy than one-vs-one on all datasets

Answer: B

Q1249. Why might a well-tuned logistic regression outperform a deep neural network for tabular data?

A. Logistic regression always converges to the global optimum unlike neural networks
B. Logistic regression can learn non-linear patterns better than neural networks
C. Deep neural networks cannot process numerical features in tabular format data
D. Deep networks need massive data and tabular datasets are often small with noise

Answer: D

Q1250. What is the effect of class imbalance on the decision boundary of a standard SVM?

A. Class imbalance has no effect on SVM decision boundaries in any scenario ever
B. The boundary shifts toward the minority class increasing its misclassification rate
C. The boundary always shifts toward the majority class regardless of data layout
D. SVM automatically handles class imbalance without any parameter adjustments

Answer: B

Q1251. In a gradient boosting ensemble, how does early stopping prevent overfitting?

A. It reduces the learning rate to exactly zero after a fixed number of trees
B. It removes the first few trees which are assumed to have the highest error
C. It monitors validation loss and stops adding trees when performance degrades
D. It limits the maximum number of features each tree can use at any split

Answer: C

Q1252. Why does XGBoost use regularization terms in its objective function unlike traditional GBDT?

A. Regularization slows training to allow more careful optimization at each step
B. It penalizes tree complexity reducing overfitting and improving generalization
C. XGBoost regularization only affects the first tree and has no effect on others
D. Regularization in XGBoost replaces the need for any learning rate parameter

Answer: B

Q1253. When would you prefer LightGBM's leaf-wise growth over XGBoost's level-wise growth?

A. When all features are categorical and no numerical features exist in data
B. When the dataset is very small and overfitting is a major concern for accuracy
C. When you need the simplest possible model with minimal hyperparameters to set
D. When training speed is critical and the dataset is large with many features

Answer: D

Q1254. How does the DART algorithm modify standard gradient boosting to reduce overfitting?

A. It randomly drops previously built trees during each boosting iteration step
B. It uses only the first and last trees and ignores all trees built in between
C. It doubles the learning rate at each iteration to converge more quickly here
D. It trains all trees simultaneously rather than sequentially to save total time

Answer: A

Q1255. In a production system, why might model ensembles be problematic despite higher accuracy?

A. Ensembles are impossible to train on any modern GPU or cloud computing platform
B. They increase inference latency, memory usage, and deployment complexity significantly
C. Ensembles always produce less accurate predictions than single model systems
D. They require manual human review for every single prediction made by the system

Answer: B

Q1256. What is the theoretical basis for why bagging reduces variance in predictions?

A. The theoretical basis is unknown and bagging works only through empirical results
B. Bagging increases variance because it uses more models than a single learner
C. Bagging reduces bias not variance since it trains on different data subsets each
D. Averaging independent estimates reduces variance by a factor of the ensemble size

Answer: D

Q1257. How does CatBoost handle categorical features differently from XGBoost and LightGBM?

A. CatBoost requires all features to be one-hot encoded before training can begin
B. CatBoost converts all categorical features into random numerical values first
C. CatBoost uses ordered target statistics to encode categoricals avoiding leakage
D. CatBoost cannot process any categorical features and only handles numerical ones

Answer: C

Q1258. In a stacking ensemble, why should base model predictions be generated using cross-validation?

A. It prevents data leakage since base models should not predict on their training data
B. It ensures that all base models have identical performance before stacking them
C. Cross-validation makes base models train faster than using the full training set
D. Cross-validation automatically selects the best base models for the final stack

Answer: A

Q1259. What is the relationship between ensemble size and the law of diminishing returns?

A. Smaller ensembles always outperform larger ensembles due to reduced complexity
B. Adding more models always linearly improves accuracy without any upper bound limit
C. After a point additional models contribute marginal gains while increasing costs
D. Ensemble size has no effect on performance and only affects training speed value

Answer: C

Q1260. Why is Bayesian optimization particularly effective for tuning ensemble hyperparameters?

A. It only works with ensemble methods and cannot tune single model parameters
B. It models the objective function to intelligently select promising configurations
C. Bayesian optimization always finds the global optimum in exactly one iteration
D. Bayesian optimization randomly samples hyperparameters like grid search does

Answer: B

Q1261. Why might K-Means converge to a suboptimal solution and how is this typically addressed?

A. Suboptimal convergence only occurs when the dataset has fewer than ten samples
B. K-Means always finds the globally optimal solution regardless of initialization
C. It can converge to local optima so K-Means++ initialization is used for better starts
D. K-Means is guaranteed to diverge without specialized GPU hardware for computation

Answer: C

Q1262. How does spectral clustering handle non-convex cluster shapes that K-Means cannot?

A. It constructs a similarity graph and uses eigenvectors of its Laplacian for clustering
B. It converts all non-convex clusters into convex ones before applying K-Means again
C. It simply increases the number of clusters K until the shapes are captured correctly
D. It uses the same distance metric as K-Means but with more iterations for accuracy

Answer: A

Q1263. What is the key advantage of UMAP over t-SNE for dimensionality reduction?

A. UMAP is a linear method while t-SNE is non-linear which makes UMAP faster here
B. UMAP requires labeled data while t-SNE works in a fully unsupervised manner only
C. UMAP better preserves global structure and scales to much larger datasets efficiently
D. UMAP always produces better visualizations than t-SNE on every possible dataset

Answer: C

Q1264. In topic modeling, how does Latent Dirichlet Allocation discover topics from documents?

A. It models documents as mixtures of topics and topics as distributions over words
B. It uses supervised labels to assign documents to predefined topic categories only
C. It requires human experts to manually define all topics before running analysis
D. It clusters documents using K-Means on word frequency vectors directly always

Answer: A

Q1265. Why is the choice of linkage criterion important in agglomerative hierarchical clustering?

A. All linkage criteria produce identical dendrograms regardless of the input dataset
B. Linkage criterion has no effect on the final clustering result in any case ever
C. Linkage criterion only affects the visualization of dendrogram and not the clusters
D. Different linkage criteria produce different cluster shapes and handle noise differently

Answer: D

Q1266. How does the Variational Autoencoder differ from a standard autoencoder for unsupervised learning?

A. Both models are identical in architecture and only differ in their training data used
B. VAE produces only deterministic encodings while standard autoencoders are probabilistic
C. VAE learns a probabilistic latent space enabling meaningful data generation sampling
D. Standard autoencoders can generate new data while VAEs can only compress existing data

Answer: C

Q1267. When evaluating clustering without ground truth labels, why is the Davies-Bouldin Index useful?

A. It can only evaluate K-Means clustering and does not work with any other algorithms
B. It always produces a score between zero and one like the accuracy metric in classification
C. It measures the ratio of within-cluster to between-cluster distances with lower being better
D. It requires ground truth labels to compute making it a supervised evaluation metric

Answer: C

Q1268. What is the fundamental limitation of using reconstruction error alone to evaluate autoencoders?

A. Low reconstruction error may indicate memorization rather than useful representation learning
B. Reconstruction error can only be computed for image data and not for tabular data
C. Reconstruction error is too computationally expensive to compute on modern hardware
D. Reconstruction error always perfectly reflects the quality of the learned latent space

Answer: A

Q1269. How does contrastive learning create useful representations without labels?

A. It uses K-Means clustering as an intermediate step before training the final model
B. It requires ground truth cluster assignments to define the contrastive loss function
C. It trains models using manually annotated labels for every single data sample used
D. It learns to pull augmented views of same samples together and push different apart

Answer: D

Q1270. Why is the Information Criterion approach preferred over the elbow method for determining optimal K?

A. Information criteria require no data and determine K from theoretical analysis alone here
B. The elbow method and information criteria always agree on the same optimal K value
C. The elbow method is always more accurate than any information criterion approach used
D. Information criteria like BIC provide a principled statistical framework for model selection

Answer: D

Q1271. In medical diagnosis, why is recall often prioritized over precision?

A. Recall is always higher than precision so it is the easier metric to optimize for
B. Missing a positive diagnosis is more dangerous than a false alarm in many cases
C. Medical models always achieve perfect precision making recall the only variable metric
D. Precision is only relevant for non-medical applications and cannot be used in health

Answer: B

Q1272. Why can nested cross-validation be necessary for unbiased model selection and evaluation?

A. Nested cross-validation is only needed for datasets with fewer than fifty observations
B. Hyperparameter tuning on the same CV folds used for evaluation causes optimistic bias
C. Standard cross-validation already provides completely unbiased model selection results
D. It is a deprecated technique that has been replaced by simple holdout validation now

Answer: B

Q1273. A model achieves 0.95 AUC but poor performance at the deployed threshold. What is the likely issue?

A. The model was trained on too much data causing it to generalize too well overall
B. The model is well-calibrated but the AUC metric was computed with an error here
C. AUC summarizes all thresholds but the operational threshold may fall in a poor region
D. An AUC of 0.95 guarantees excellent performance at every possible threshold value

Answer: C

Q1274. When comparing models, why is statistical significance testing important beyond comparing mean metrics?

A. Statistical testing is only applicable to models trained on datasets exceeding millions
B. Performance differences may be due to random variation rather than true model superiority
C. Mean metrics always provide sufficient information to select the best model reliably
D. It is never needed because cross-validation already accounts for all statistical variance

Answer: B

Q1275. What is the Matthews Correlation Coefficient and why is it preferred for imbalanced datasets?

A. It is identical to accuracy and provides no additional information for imbalanced data
B. It only considers true positives and ignores all other confusion matrix cell values
C. It uses all confusion matrix values providing balanced evaluation even with skewed classes
D. It requires balanced classes to compute and returns undefined for imbalanced datasets

Answer: C

Q1276. How does the precision-recall curve complement the ROC curve for rare event detection?

A. Both curves always provide identical information regardless of the class balance ratio
B. PR curves better expose model performance on the minority class that ROC may obscure
C. ROC curves are always superior to PR curves in every evaluation scenario encountered
D. PR curves can only be computed when classes are perfectly balanced in the dataset

Answer: B

Q1277. Why might a model perform well in cross-validation but poorly in real-world deployment?

A. Distribution shift between training data and production data degrades model performance
B. Cross-validation uses test data for training which inflates deployment expectations
C. Models that perform well in CV are guaranteed to perform well in any deployment setup
D. Cross-validation always provides perfectly accurate estimates of deployment performance

Answer: A

Q1278. What is the purpose of the DeLong test in comparing ROC curves of two models?

A. It statistically tests whether two AUC values are significantly different from each other
B. It measures the calibration quality of both models probability estimates simultaneously
C. It determines whether one model trains faster than the other on the same dataset
D. It automatically selects the optimal threshold for deploying both models in production

Answer: A

Q1279. In time-series model evaluation, why is standard K-fold cross-validation inappropriate?

A. Standard K-fold is actually the recommended approach for all time-series evaluations
B. Random fold assignment violates temporal ordering and leaks future data into training
C. Time-series data always has too few samples for any cross-validation method to work
D. K-fold cross-validation can only be applied to classification not regression problems

Answer: B

Q1280. How does Bayesian evaluation of classifiers improve upon frequentist hypothesis testing?

A. It provides probability distributions over performance differences enabling richer analysis
B. Bayesian evaluation only works with neural network classifiers and no other methods
C. It eliminates the need for any test data by computing performance from prior alone
D. Bayesian methods always agree with frequentist tests making them completely redundant

Answer: A

Q1281. Why can target encoding cause data leakage and how is it typically mitigated?

A. It uses target values during encoding so leave-one-out or fold-based encoding prevents leak
B. Target encoding never causes data leakage so no mitigation is ever required
C. Data leakage only occurs with one-hot encoding and never with target encoding method
D. Target encoding automatically prevents leakage through its internal smoothing mechanism

Answer: A

Q1282. When engineering features for a gradient boosted tree model, why is binning continuous features usually unnecessary?

A. Tree-based models already find optimal split points so binning loses useful granularity
B. Gradient boosted trees cannot process binned features and require raw continuous input
C. Binning always improves gradient boosted tree performance without any exceptions
D. Binning is only useful for neural networks and cannot be applied to any tree models

Answer: A

Q1283. How does autoML automate feature engineering and what are its limitations?

A. AutoML perfectly replaces all manual feature engineering without any limitations ever
B. AutoML only works with image data and cannot handle tabular or text feature engineering
C. It explores transformations automatically but may miss domain-specific meaningful features
D. It always produces fewer features than manual engineering making it strictly inferior

Answer: C

Q1284. Why might feature engineering differ between linear models and tree-based models?

A. Tree-based models require normalized features while linear models handle raw values
B. Both model types benefit from exactly identical feature engineering approaches always
C. Linear models need non-linear transformations while trees handle non-linearity natively
D. Feature engineering is only needed for tree-based models and never for linear models

Answer: C

Q1285. What is the risk of using highly engineered features that are specific to the training distribution?

A. Engineered features are always more robust than raw features in all scenarios tested
B. The risk is minimal because feature engineering always reduces overfitting in models
C. They may not generalize to production data if the distribution shifts over time period
D. Highly engineered features always generalize perfectly to any future data distribution

Answer: C

Q1286. How does mutual information differ from Pearson correlation for feature selection?

A. Pearson captures non-linear relationships while mutual information is limited to linear
B. Both methods measure identical relationships and always produce the same feature ranking
C. Mutual information only works with binary features while Pearson works with all types
D. Mutual information captures any dependency while Pearson only detects linear relationships

Answer: D

Q1287. In NLP, why might TF-IDF features outperform raw word count features for text classification?

A. Raw word counts always outperform TF-IDF so this premise is incorrect in all cases
B. TF-IDF downweights common words and upweights distinctive words improving discrimination
C. TF-IDF converts words to their synonyms before counting improving vocabulary coverage
D. TF-IDF always produces fewer features than raw word counts making models train faster

Answer: B

Q1288. What is the purpose of creating lag features in time-series feature engineering?

A. They capture temporal dependencies by using past values as features for current predictions
B. Lag features convert time-series data into image format for convolutional neural networks
C. They eliminate seasonality from time-series data making it stationary for all models
D. Lag features remove temporal dependencies to make data suitable for standard ML models

Answer: A

Q1289. How does the permutation importance method evaluate feature relevance after model training?

A. It sorts features alphabetically and assigns importance based on their position order
B. It shuffles each feature's values and measures how much model performance degrades
C. It adds random noise to each feature and measures how much training accuracy improves
D. It removes each feature entirely and retrains the model from scratch to compare results

Answer: B

Q1290. Why should you be cautious about using future information when engineering features for time-series models?

A. Using future values creates temporal leakage leading to unrealistically optimistic evaluation
B. Future information is always available at prediction time so there is no risk in using it
C. Future information always improves time-series model accuracy without any negative effects
D. Time-series models automatically detect and exclude future information during training

Answer: A

Q1291. Why do residual connections in ResNet help train very deep networks effectively?

A. They allow gradients to flow directly through skip connections preventing vanishing
B. Residual connections double the number of parameters in each layer of network
C. Residual connections remove the need for any activation functions in the network
D. They automatically reduce the learning rate as the network depth increases here

Answer: A

Q1292. What is the dying ReLU problem and how is Leaky ReLU a solution?

A. Dying ReLU only affects the output layer while Leaky ReLU only affects hidden layers
B. ReLU never has this problem and Leaky ReLU was created for a different purpose
C. Dying ReLU means the function grows too large while Leaky ReLU caps the output
D. Neurons with negative inputs permanently output zero and Leaky ReLU allows small negatives

Answer: D

Q1293. How does the attention mechanism in neural networks improve sequence modeling?

A. It forces the model to process all elements equally regardless of their relevance
B. It removes the need for any training data by using pre-computed attention weights
C. It limits the model to processing only the first and last elements of sequences
D. It allows the model to dynamically focus on relevant parts of the input sequence

Answer: D

Q1294. Why is the choice of weight initialization strategy critical for deep network training?

A. Weight initialization has no effect on training since optimizers always find the same solution
B. Weight initialization only matters for the first layer and has no effect on deeper layers
C. Poor initialization can cause vanishing or exploding activations preventing effective learning
D. All initialization strategies produce identical results after sufficient training epochs pass

Answer: C

Q1295. What is the lottery ticket hypothesis and its implications for neural network pruning?

A. It suggests that random network architectures always outperform carefully designed ones
B. It proves that larger networks always outperform smaller networks without any exceptions
C. Dense networks contain sparse subnetworks that achieve comparable accuracy when trained alone
D. It states that all neural network weights are equally important and none can be removed

Answer: C

Q1296. How does knowledge distillation compress a large model into a smaller one?

A. It copies all weights directly from the large model into the smaller model architecture
B. It trains the small model only on data that the large model classified incorrectly
C. A smaller student model learns to mimic the soft output probabilities of a larger teacher
D. It removes random layers from the large model until the desired size is reached

Answer: C

Q1297. Why do neural architecture search methods face a massive computational challenge?

A. The challenge is purely theoretical and modern hardware makes NAS trivially inexpensive
B. NAS only searches over learning rates not architectures so the space is very small
C. The search space of possible architectures is exponentially large requiring many evaluations
D. Neural architecture search always converges in a single iteration without much computation

Answer: C

Q1298. What role does the temperature parameter play in softmax for knowledge distillation?

A. Temperature only affects training speed and has no effect on the output distribution
B. Temperature is fixed at one and cannot be changed in any softmax implementation ever
C. Lower temperature always produces better distillation results regardless of the task
D. Higher temperature produces softer probabilities revealing more about learned similarities

Answer: D

Q1299. How does mixed precision training accelerate deep learning while maintaining accuracy?

A. It halves the training data to reduce computation time at the cost of some accuracy
B. It skips every other layer during forward pass to achieve faster computation speed
C. It uses float16 for most operations and float32 for critical ones reducing memory and time
D. It uses only 8-bit integers for all computations throughout the entire training process

Answer: C

Q1300. What is the neural tangent kernel framework and why is it theoretically important?

A. It is a specific type of convolutional kernel used only for image recognition tasks
B. It is a pruning technique for removing unnecessary connections from trained networks
C. It is an optimizer that replaces gradient descent for all neural network training
D. It shows infinite-width networks behave like kernel methods enabling theoretical analysis

Answer: D

Q1301. Why do Vision Transformers require large-scale pre-training to outperform CNNs?

A. Vision Transformers have fewer parameters than CNNs and need more data to compensate
B. CNNs always outperform Vision Transformers regardless of the training data scale used
C. Vision Transformers can only process text data and need pre-training to handle images
D. ViTs lack inductive biases like locality and translation equivariance that CNNs have

Answer: D

Q1302. How does the Wasserstein loss improve GAN training stability compared to the original formulation?

A. It eliminates the need for a discriminator network in the adversarial training setup
B. It provides smoother gradients based on earth mover distance even when distributions differ
C. Wasserstein loss only works with image generation and not with other data modalities
D. It uses the same loss as the original GAN but with a different optimizer algorithm

Answer: B

Q1303. What is the computational complexity of self-attention and how do efficient variants address it?

A. Efficient variants make attention slower but produce more accurate attention weights
B. Self-attention is O(n) and efficient variants increase it to O(n squared) for accuracy
C. Self-attention is O(n squared) and variants like Linformer use projections for O(n)
D. Self-attention has constant complexity regardless of the sequence length provided

Answer: C

Q1304. How does the diffusion model generate high-quality images compared to GANs?

A. They require significantly less training data than GANs to produce comparable results
B. Diffusion models train adversarially like GANs but with more discriminator networks
C. Diffusion models generate images in a single forward pass without any iterative steps
D. They learn to reverse a gradual noising process producing diverse high-quality samples

Answer: D

Q1305. What is the significance of the universal approximation theorem for neural networks?

A. It guarantees that deeper networks always outperform shallower networks on all tasks
B. A sufficiently wide single-layer network can approximate any continuous function closely
C. It states that neural networks can only approximate linear functions and nothing else
D. It proves that neural networks always find the global optimum during training runs

Answer: B

Q1306. How does neural architecture search with weight sharing reduce computational cost?

A. It trains each candidate architecture from scratch independently on separate hardware
B. It searches only over learning rates not architectures making the space much smaller
C. Multiple architectures share weights from a supernet avoiding full retraining each time
D. Weight sharing eliminates the need for any gradient computation during architecture search

Answer: C

Q1307. Why is training stability a major challenge in GAN optimization?

A. GAN training is always stable and converges to the optimal solution every single time
B. The minimax game between generator and discriminator can oscillate without convergence
C. The generator and discriminator always converge at the exact same rate making it easy
D. Stability issues only occur with very small datasets of fewer than hundred total samples

Answer: B

Q1308. How does the mixture of experts architecture achieve computational efficiency in large models?

A. It reduces model size by removing most parameters from each expert subnetwork
B. It uses a single large expert for all inputs and ignores the routing mechanism
C. It routes each input to only a subset of experts reducing computation per sample
D. It trains all expert networks on every input sample to ensure comprehensive coverage

Answer: C

Q1309. What is the role of contrastive loss in self-supervised representation learning?

A. It replaces all other loss functions and is the only one used in deep learning
B. It only works with generative models and cannot be used with discriminative ones
C. It pulls similar sample representations together and pushes dissimilar ones apart
D. It requires labeled data pairs to compute the loss function during training phase

Answer: C

Q1310. How does the Flash Attention algorithm improve Transformer training efficiency?

A. It uses tiling and recomputation to reduce memory IO making attention faster on GPU
B. Flash Attention only works on CPU hardware and provides no benefit on GPU systems
C. It reduces attention accuracy to save computation time during the training process
D. It removes the attention mechanism entirely replacing it with simple averaging only

Answer: A

Q1311. Why do large language models sometimes generate factually incorrect but plausible-sounding text?

A. LLMs verify all facts against a database before generating any output text content
B. LLMs always generate perfectly accurate text since they are trained on factual data
C. Factual errors only occur when the model has fewer than one million total parameters
D. LLMs learn statistical patterns rather than factual knowledge causing hallucination errors

Answer: D

Q1312. How does the Transformer's multi-head attention provide an advantage over single-head attention?

A. Multi-head attention is identical to single-head attention and provides no real advantage
B. Multiple heads only work with image data and provide no benefit for text processing
C. Multiple heads attend to different representation subspaces capturing diverse relationships
D. Multi-head attention uses fewer parameters than single-head attention in all cases

Answer: C

Q1313. What is the fundamental challenge of evaluating open-ended text generation models?

A. Open-ended generation models always produce identical outputs for the same input text
B. Evaluation is trivial because BLEU score perfectly captures all aspects of text quality
C. There is always exactly one correct output for any given text generation prompt used
D. Multiple valid outputs exist making automatic metrics insufficient for quality assessment

Answer: D

Q1314. How does reinforcement learning from human feedback improve language model alignment?

A. RLHF only affects the tokenizer and has no impact on the language model weights
B. It fine-tunes the model using a reward signal trained on human preference judgments
C. It replaces the Transformer architecture with a recurrent neural network approach
D. RLHF removes all pre-training knowledge and trains the model entirely from scratch

Answer: B

Q1315. Why is the position of negation words critical for sentiment analysis accuracy?

A. Negation words have no effect on sentiment and can be safely removed during preprocessing
B. Negation reverses sentiment of subsequent words and mishandling it causes incorrect polarity
C. Negation only appears in formal text and never occurs in informal social media content
D. All negation words are stop words and are automatically removed by all NLP pipelines used

Answer: B

Q1316. What is the catastrophic forgetting problem when fine-tuning pre-trained language models?

A. Catastrophic forgetting only occurs in computer vision models and never in NLP models
B. The model loses general pre-trained knowledge when heavily adapted to a specific new task
C. Fine-tuning always improves performance on all tasks simultaneously without any tradeoffs
D. Fine-tuning adds new knowledge without affecting any previously learned representations

Answer: B

Q1317. How do retrieval-augmented generation models reduce hallucination in language models?

A. They remove the generative component entirely and only return retrieved document text
B. They retrieve relevant documents and condition generation on factual source material
C. RAG models eliminate hallucination completely with a one hundred percent success rate
D. They increase the model size to store more factual information in the network parameters

Answer: B

Q1318. Why is cross-lingual transfer learning effective for low-resource languages?

A. Multilingual models learn shared representations that transfer knowledge between languages
B. Cross-lingual transfer degrades performance on all languages including high-resource ones
C. Each language requires a completely separate model with no knowledge sharing possible
D. Cross-lingual transfer only works between languages that share the same script system

Answer: A

Q1319. What is the computational challenge of scaling attention to very long document contexts?

A. Long documents always produce better results so there is no computational challenge here
B. Long documents can be processed in constant time regardless of their total length overall
C. Attention memory and compute scale quadratically with sequence length becoming prohibitive
D. Attention complexity is independent of sequence length and depends only on model depth

Answer: C

Q1320. How does chain-of-thought prompting improve reasoning in large language models?

A. It breaks complex problems into intermediate steps allowing systematic reasoning output
B. It reduces the model's accuracy but makes the output more readable for human reviewers
C. Chain-of-thought only works for arithmetic and cannot improve any other reasoning tasks
D. It forces the model to generate the answer in a single token without any explanation

Answer: A

Q1321. Why do adversarial examples pose a serious threat to deployed computer vision systems?

A. Adversarial examples only affect models during training and never at inference time
B. All modern CNN architectures are completely immune to any adversarial perturbations
C. Adversarial examples only exist in theory and cannot be generated for real images
D. Imperceptible perturbations can cause confident misclassification in production models

Answer: D

Q1322. How does contrastive learning for visual representations reduce dependence on labeled data?

A. It completely replaces supervised fine-tuning and achieves perfect accuracy on all tasks
B. Contrastive learning only works with text data and cannot be applied to images ever
C. It requires more labeled data than supervised learning to achieve comparable accuracy
D. It learns representations by contrasting augmented views of images without any labels

Answer: D

Q1323. What is the domain adaptation challenge when deploying CV models across different environments?

A. Distribution differences between training and deployment environments degrade performance
B. Domain adaptation is only relevant for NLP tasks and does not apply to computer vision
C. All visual environments produce identical image statistics so adaptation is unnecessary
D. Models trained in one visual domain always generalize perfectly to all other domains

Answer: A

Q1324. How does the panoptic segmentation task unify semantic and instance segmentation?

A. It assigns every pixel both a semantic class label and an instance identifier value
B. It replaces both tasks with a simpler bounding box detection approach for efficiency
C. Panoptic segmentation only works on outdoor scenes and cannot process indoor images
D. It only performs semantic segmentation and completely ignores instance-level information

Answer: A

Q1325. Why is 3D object detection from point clouds fundamentally different from 2D image detection?

A. Point clouds are identical to images and can be processed by standard 2D CNN methods
B. 3D detection is always easier than 2D detection because more information is available
C. Point clouds are sparse, unordered, and require specialized architectures like PointNet
D. Point clouds can only represent indoor environments and not outdoor driving scenarios

Answer: C

Q1326. What is the role of the feature extractor backbone in modern object detection frameworks?

A. It handles data augmentation and preprocessing before the actual detection begins
B. It generates the final bounding box predictions and class labels for each object found
C. The backbone is only used during training and is removed completely during inference
D. It extracts hierarchical visual features that the detection head uses for predictions

Answer: D

Q1327. How does self-supervised pre-training with masked image modeling work for vision models?

A. It only works with text-image pairs and cannot be applied to images alone at all
B. It masks random image patches and trains the model to reconstruct the missing content
C. It requires fully labeled segmentation maps for every image in the training dataset
D. It trains models to classify images into predefined categories using labeled datasets

Answer: B

Q1328. Why is temporal consistency important in video object detection and how is it maintained?

A. Temporal consistency is irrelevant because each video frame should be processed independently
B. Video frames are always identical so temporal consistency is automatically guaranteed here
C. Flickering predictions between frames indicate instability and tracking methods enforce consistency
D. Temporal consistency only matters for video compression and has no effect on detection

Answer: C

Q1329. What is the advantage of vision-language models like CLIP for zero-shot image classification?

A. They require separate fine-tuning for every possible class label before making predictions
B. Vision-language models can only classify images into the exact categories seen during training
C. They always perform worse than supervised models regardless of the number of training classes
D. They match images to text descriptions enabling classification without task-specific training

Answer: D

Q1330. How does neural radiance field technology enable novel view synthesis from images?

A. It requires thousands of input images from every possible angle of the target scene
B. It simply stitches existing images together without any three-dimensional understanding
C. It learns a volumetric scene representation mapping 3D coordinates to color and density
D. NeRF only works with synthetic computer-generated scenes and not real-world photographs

Answer: C

Q1331. Why is exactly-once processing semantics challenging to achieve in distributed streaming systems?

A. Network failures and node crashes can cause duplicates or losses requiring complex protocols
B. Exactly-once processing is trivial and all modern streaming systems implement it easily
C. Streaming systems never experience failures so exactly-once is guaranteed automatically
D. Exactly-once only matters for batch processing and is irrelevant for streaming systems

Answer: A

Q1332. How does the Lambda architecture address the needs of both batch and real-time analytics?

A. Lambda architecture only supports batch processing and has no real-time component
B. It maintains separate batch and speed layers that serve a combined serving layer
C. It uses only a single processing layer for both batch and real-time data analysis
D. It replaces all batch processing with stream processing for unified data handling

Answer: B

Q1333. What is data skew and why does it degrade performance in distributed processing?

A. Data skew only affects storage and has no impact on query processing performance
B. Data skew always improves performance by concentrating work on the fastest machines
C. All distributed systems automatically handle data skew without any intervention
D. Uneven data distribution causes some nodes to be overloaded while others sit idle

Answer: D

Q1334. How does the Kappa architecture simplify the Lambda architecture?

A. Kappa uses only a stream processing layer treating everything as streams eliminating batch
B. Kappa removes the speed layer and relies solely on batch processing for all analytics
C. Kappa adds a third processing layer on top of Lambda's existing batch and speed layers
D. Kappa and Lambda architectures are identical and the terms are completely interchangeable

Answer: A

Q1335. Why is shuffle operation one of the most expensive operations in distributed data processing?

A. Shuffle operations only move data within a single machine which is very fast operation
B. Shuffles only occur in batch processing and never happen in stream processing pipelines
C. Shuffles move data across the network between nodes causing significant IO and latency
D. Shuffle operations are optimized by all frameworks and never cause performance issues

Answer: C

Q1336. How does delta lake improve upon traditional data lake architectures?

A. It replaces the distributed file system with a single centralized database for storage
B. Delta lake removes all structure and stores data as completely unformatted text files
C. It adds ACID transactions and schema enforcement to data lake storage for reliability
D. Delta lake only works with structured data and cannot store any unstructured data files

Answer: C

Q1337. What are the tradeoffs between using a wide table versus a normalized schema in big data?

A. There are no tradeoffs as both approaches produce identical query performance results
B. Wide tables are always superior to normalized schemas for every possible query pattern
C. Wide tables reduce joins but increase redundancy while normalized schemas minimize redundancy
D. Normalized schemas always perform better than wide tables in distributed query engines

Answer: C

Q1338. How does approximate query processing enable faster analytics on massive datasets?

A. It only works for count queries and cannot approximate any other aggregation types
B. It produces exact results faster by using more efficient algorithms on all the data present
C. Approximate processing always produces completely wrong results with no useful insights
D. It trades some accuracy for speed using sampling and sketches on subsets of the data

Answer: D

Q1339. Why is data lineage tracking important in complex big data pipelines?

A. Data lineage is only useful for documentation and has no practical operational value
B. It traces data transformations enabling debugging, compliance, and impact analysis
C. Data lineage tracking always degrades pipeline performance by more than fifty percent
D. It only applies to batch processing and cannot track streaming data transformations

Answer: B

Q1340. How does federated learning enable ML on distributed data without centralizing it?

A. It only works with image data and cannot be applied to tabular or text data formats
B. Models train locally on each node and only share model updates not raw data for privacy
C. Federated learning always produces worse models than centralized training on all tasks
D. It requires all data to be collected in a central location before any training can begin

Answer: B

Q1341. How does a feature store ensure consistency between training and serving in production ML?

A. Feature stores only handle training features and have no role in serving predictions
B. It provides a single source of truth for feature definitions and computation logic used
C. Feature stores always cause training-serving skew due to different code paths in them
D. It stores raw data only and requires separate feature computation for training and serving

Answer: B

Q1342. Why is reproducibility challenging in ML systems and how does MLOps address it?

A. ML systems are perfectly reproducible by default and require no additional tooling at all
B. MLOps makes all models deterministic by fixing random seeds and nothing else is needed
C. Reproducibility is only important in academic research and irrelevant for production ML
D. Non-determinism in training, data changes, and environment drift require systematic tracking

Answer: D

Q1343. What is the role of a model registry in enterprise MLOps workflows?

A. Model registries are only useful for small teams and provide no value at enterprise scale
B. It automatically trains new models without any human oversight or approval workflows
C. It manages model lifecycle including versioning, staging, approval, and deployment tracking
D. It only stores model code and has no capability to track model metadata or lineage

Answer: C

Q1344. How does data-centric AI differ from model-centric AI in improving ML system performance?

A. Data-centric AI only applies to unstructured data and model-centric to structured data
B. Data-centric and model-centric approaches are identical and interchangeable in all cases
C. Data-centric focuses on improving data quality while model-centric focuses on architecture
D. Model-centric AI never requires any training data and relies solely on model architecture

Answer: C

Q1345. What are the challenges of implementing real-time ML inference at scale?

A. Real-time inference is trivial to implement and never poses any engineering challenges
B. All models inherently provide sub-millisecond inference without any optimization needed
C. Latency requirements, throughput scaling, and model optimization create significant complexity
D. Real-time inference only works with batch predictions and cannot handle individual requests

Answer: C

Q1346. How does GitOps extend traditional DevOps practices for ML workflow management?

A. It eliminates the need for any CI/CD pipelines by manually deploying all changes to prod
B. It uses Git as the single source of truth for declarative infrastructure and ML pipeline config
C. GitOps replaces Git with a custom version control system designed only for ML artifacts
D. GitOps only works with data preprocessing and cannot manage model training or deployment

Answer: B

Q1347. Why is testing ML systems more complex than testing traditional software?

A. Testing ML systems requires only standard unit tests identical to traditional software ones
B. ML systems are simpler to test because they only require checking if the code compiles
C. ML involves data dependencies, model behavior, and performance thresholds beyond unit tests
D. ML systems never need testing because model accuracy guarantees correct system behavior

Answer: C

Q1348. What is the purpose of a model serving framework like TensorFlow Serving or Triton?

A. They replace the need for any monitoring or logging of deployed model performance metrics
B. They optimize and serve models efficiently handling batching, versioning, and concurrency
C. They are only used for training models and have no capability for serving predictions
D. Model serving frameworks only work with a single model and cannot manage multiple models

Answer: B

Q1349. How does ML governance address regulatory requirements for AI systems in production?

A. Governance eliminates the need for any model monitoring after initial deployment is done
B. It only applies to models making financial decisions and not to any other domain areas
C. It provides audit trails, documentation, fairness monitoring, and compliance frameworks
D. ML governance is only relevant for academic research and has no production implications

Answer: C

Q1350. What is the concept of ML technical debt and how does it accumulate in production systems?

A. Technical debt only applies to traditional software and is irrelevant for ML systems always
B. ML technical debt is automatically resolved by retraining the model on newer data only
C. ML systems never accumulate technical debt because they are self-maintaining by design
D. Data dependencies, pipeline complexity, and configuration debt compound over time in systems

Answer: D

Q1351. Why is it mathematically impossible to satisfy all fairness criteria simultaneously in most real-world scenarios?

A. All fairness criteria can be satisfied simultaneously with enough training data available
B. Different fairness metrics are fundamentally incompatible when base rates differ across groups
C. Only linear models face this limitation while neural networks can satisfy all criteria easily
D. Impossibility only applies to theoretical scenarios and never to practical AI applications

Answer: B

Q1352. How does federated learning attempt to address privacy concerns while enabling collaborative ML?

A. It trains models locally sharing only model updates while keeping raw data decentralized
B. It requires centralizing all data from all participants before any model training begins
C. Federated learning provides no privacy benefits compared to centralized training approaches
D. It encrypts all data using standard AES encryption and sends it to a central server only

Answer: A

Q1353. What is the value alignment problem in the context of advanced AI systems?

A. Ensuring AI systems pursue goals consistent with complex and sometimes conflicting human values
B. Value alignment is only relevant for recommendation systems and not for other AI applications
C. It refers to aligning model weights with optimal values during the training optimization
D. It guarantees AI systems will always act in humanity's best interest without any intervention

Answer: A

Q1354. How can model cards and datasheets improve responsible AI development practices?

A. They replace the need for any testing or evaluation of AI models before deployment occurs
B. Model cards are only useful for marketing purposes and provide no technical or ethical value
C. They provide structured documentation about model capabilities, limitations, and intended use
D. They automatically fix all bias issues in AI models without any human intervention needed

Answer: C

Q1355. What is the tension between model accuracy and fairness in high-stakes AI applications?

A. Higher accuracy always guarantees better fairness outcomes for all groups simultaneously
B. Optimizing for fairness constraints may reduce accuracy and vice versa requiring tradeoffs
C. There is never any tension because accurate models are always fair to all groups by nature
D. Fairness constraints always improve model accuracy making the tension nonexistent overall

Answer: B

Q1356. How do deepfakes pose unique challenges for AI ethics and society?

A. They enable realistic fabrication of content undermining trust in media and enabling fraud
B. Deepfakes are easily detectable by all humans and pose no significant societal challenges
C. Deepfakes are only used for entertainment purposes and have no potential for misuse ever
D. Deepfake technology can only produce low-quality content that no one would believe is real

Answer: A

Q1357. Why is the concept of meaningful human oversight important for autonomous AI systems?

A. Meaningful oversight only applies to AI systems in military applications and no other fields
B. Humans must retain ability to understand, intervene, and override AI decisions when needed
C. Human oversight should be completely removed to maximize the efficiency of AI systems now
D. Autonomous AI systems never make errors so human oversight is unnecessary and wasteful

Answer: B

Q1358. What are the environmental concerns associated with training large AI models?

A. Training large AI models has no environmental impact and uses negligible energy resources
B. Large model training consumes significant energy and computing resources increasing carbon footprint
C. Environmental concerns only apply to models trained on renewable energy sources specifically
D. Training small models produces more emissions than large models due to lower computational efficiency

Answer: B

Q1359. How does the EU AI Act categorize AI systems based on risk levels?

A. The EU AI Act bans all AI systems regardless of their risk level or application domain
B. The EU AI Act only applies to military AI systems and exempts all commercial applications
C. It treats all AI systems identically without any risk-based categorization or differentiation
D. It classifies systems into unacceptable, high, limited, and minimal risk with corresponding rules

Answer: D

Q1360. What is the challenge of ensuring AI accountability when AI systems make autonomous decisions?

A. AI systems can be held legally accountable themselves without any human responsibility needed
B. Accountability is only relevant when AI causes physical harm and not for any other type of impact
C. AI accountability is straightforward since the developer is always clearly responsible for everything
D. Complex AI systems make it difficult to attribute responsibility across developers, deployers, and users

Answer: D

Q1361. What is the Moravec paradox in AI research?

A. Simple computational tasks are harder for AI than complex reasoning
B. High-level reasoning requires less computation than sensorimotor skills
C. AI systems excel at tasks humans find easy but struggle with tasks humans find hard
D. Tasks easy for humans like perception and mobility are extremely hard for AI while abstract reasoning is comparatively easier

Answer: D

Q1362. What is the difference between deductive and inductive reasoning in AI systems?

A. They are identical approaches
B. Deductive reasoning goes from general rules to specific conclusions while inductive reasoning infers general rules from specific observations
C. Deductive reasoning is always better than inductive
D. Inductive reasoning does not use data

Answer: B

Q1363. What is the computational theory of mind and how does it relate to AI?

A. Computers are conscious beings
B. The mind operates like a computational system processing information, providing theoretical basis for AI
C. All computations are intelligent
D. Minds cannot be modeled computationally

Answer: B

Q1364. What is the Bitter Lesson identified by Rich Sutton in AI research?

A. AI research is always unsuccessful
B. General methods leveraging computation scale better than methods leveraging human domain knowledge
C. Human knowledge is always superior to machine learning
D. Smaller models always outperform larger ones

Answer: B

Q1365. What is the concept of emergence in complex AI systems?

A. AI systems always behave predictably
B. Complex behaviors and capabilities arising from simpler components that were not explicitly programmed
C. All AI capabilities must be manually coded
D. Emergence only occurs in biological systems

Answer: B

Q1366. How does the concept of bounded rationality apply to AI agent design?

A. AI agents always have unlimited resources
B. Agents must make satisficing decisions due to limited computational resources and information
C. Bounded rationality means AI cannot learn
D. It only applies to human decision-making

Answer: B

Q1367. What is the difference between symbolic AI and connectionist AI?

A. They are the same approach
B. Symbolic AI uses explicit rules and logic while connectionist AI uses neural networks to learn representations
C. Symbolic AI is always superior
D. Connectionist AI does not use mathematics

Answer: B

Q1368. What is the scaling hypothesis in modern AI research?

A. Smaller models are always better
B. Increasing model size, data, and compute will continue to yield improved capabilities and potentially lead to more general intelligence
C. Scaling always leads to diminishing returns
D. AI models cannot be scaled beyond a fixed size

Answer: B

Q1369. What is the concept of an AI winter and what historically caused them?

A. A season when AI labs close for vacation
B. Periods of reduced funding and interest in AI due to unmet expectations and hype cycles
C. AI systems that only work in cold temperatures
D. A type of cooling system for AI hardware

Answer: B

Q1370. What is multi-task learning and why can it improve generalization?

A. Training separate models for each task independently
B. Training a single model on multiple related tasks simultaneously so shared representations improve performance on each task
C. It always reduces model accuracy
D. Multi-task learning only works for image tasks

Answer: B

Q1371. Why is the softmax function a generalization of the sigmoid function for multiple classes?

A. They are completely unrelated functions
B. Softmax reduces to sigmoid for the two-class case and extends the logistic function to produce a probability distribution over K classes
C. Sigmoid is more general than softmax
D. Softmax only works with two classes

Answer: B

Q1372. What is the connection between maximum likelihood estimation and cross-entropy loss?

A. They are unrelated concepts
B. Minimizing cross-entropy loss is mathematically equivalent to maximizing the log-likelihood of the data under the model
C. Cross-entropy always equals zero
D. MLE does not apply to classification

Answer: B

Q1373. What is the significance of the spectral norm of a matrix in deep learning?

A. It only applies to image processing
B. The largest singular value that controls the Lipschitz constant and affects training stability in neural networks
C. It is the sum of all matrix elements
D. It measures matrix sparsity

Answer: B

Q1374. How does the reparameterization trick enable backpropagation through stochastic nodes?

A. It removes all randomness from the model
B. It separates the stochastic component from learnable parameters so gradients can flow through the deterministic path
C. It doubles the learning rate
D. It is only used in supervised learning

Answer: B

Q1375. What is the role of the Lagrangian in constrained optimization for ML?

A. It is a type of neural network layer
B. It converts constrained optimization into an unconstrained problem by incorporating constraints as penalty terms with multipliers
C. It only applies to linear regression
D. It eliminates the need for gradient descent

Answer: B

Q1376. Why is the positive semi-definiteness of a kernel matrix required in kernel methods?

A. It makes computation faster
B. It ensures the kernel implicitly computes inner products in a valid feature space per Mercer's theorem
C. It is not actually required
D. It guarantees the matrix is invertible

Answer: B

Q1377. What is the information-theoretic interpretation of entropy in machine learning?

A. It measures dataset size
B. It quantifies the expected amount of information or uncertainty in a random variable's outcomes
C. It is the same as variance
D. It only applies to text data

Answer: B

Q1378. How does the matrix rank relate to the information content of a dataset?

A. Rank always equals the number of rows
B. The rank indicates the number of linearly independent dimensions, revealing the true dimensionality and potential redundancy in the data
C. Rank is always equal to one
D. It measures the speed of matrix operations

Answer: B

Q1379. What is the relationship between the trace of a matrix and eigenvalues?

A. They are unrelated
B. The trace equals the sum of eigenvalues, connecting matrix operations to spectral properties
C. The trace is always zero
D. Eigenvalues cannot be summed

Answer: B

Q1380. Why does stochastic gradient descent converge despite using noisy gradient estimates?

A. It does not actually converge
B. Under conditions of decreasing learning rates and bounded variance, the noise averages out and SGD converges to a local minimum
C. SGD always finds the global minimum
D. Noise prevents any form of convergence

Answer: B

Q1381. How does Python's garbage collector handle circular references?

A. It cannot handle them at all
B. It uses a generational garbage collector that detects and collects reference cycles beyond simple reference counting
C. Circular references never occur in Python
D. It requires manual memory management

Answer: B

Q1382. What is the purpose of __init__.py in Python packages?

A. It initializes variables to zero
B. It marks a directory as a Python package, can run initialization code, and controls what is exported via __all__
C. It is required for every Python file
D. It only contains comments

Answer: B

Q1383. Why is joblib preferred over pickle for serializing scikit-learn models with large NumPy arrays?

A. Joblib is newer and therefore better
B. Joblib efficiently handles large NumPy arrays by using memory mapping and optimized serialization for numerical data
C. Pickle cannot serialize any ML models
D. Joblib produces smaller files for all data types

Answer: B

Q1384. What is the difference between asyncio and multiprocessing for concurrent ML workloads?

A. They are identical approaches
B. asyncio handles I/O-bound concurrency in a single thread while multiprocessing uses separate processes for CPU-bound parallelism
C. asyncio is always faster
D. multiprocessing cannot be used for ML

Answer: B

Q1385. How do Python descriptors work and where are they used in ML frameworks?

A. They describe variable names
B. Descriptors are objects defining __get__, __set__, or __delete__ methods that control attribute access, used in frameworks for parameter validation and lazy loading
C. They are only used for printing
D. Descriptors are a type of loop

Answer: B

Q1386. What are metaclasses in Python and how do ML frameworks use them?

A. They are classes that define how other classes are created, used in frameworks for automatic registration and validation of model components
B. They are a type of data structure
C. Metaclasses only exist in Java
D. They are the same as abstract classes

Answer: A

Q1387. Why is using numpy.random.Generator preferred over the legacy numpy.random functions?

A. It produces the same results
B. Generator provides better statistical properties, reproducibility via explicit seed management, and thread safety compared to the global random state
C. Legacy functions are faster
D. There is no difference in practice

Answer: B

Q1388. How does Cython improve performance for computationally intensive ML code?

A. It replaces Python entirely
B. It compiles Python-like code to C, enabling static type declarations and direct C-level operations that bypass Python interpreter overhead
C. Cython only works for web applications
D. It automatically parallelizes all code

Answer: B

Q1389. What is the purpose of the __repr__ and __str__ methods in custom ML classes?

A. They delete the object
B. __repr__ provides an unambiguous developer-facing representation while __str__ provides a user-friendly readable string
C. They are identical methods
D. They are used for encryption

Answer: B

Q1390. What is the advantage of using Python's dataclasses for ML configuration management?

A. They are faster than regular classes
B. They auto-generate __init__, __repr__, __eq__ and support type hints, reducing boilerplate for configuration objects while maintaining readability
C. They cannot store numerical data
D. They replace all other class types

Answer: B

Q1391. How does the iterative imputer (MICE) handle missing data differently from simple imputation?

A. It deletes rows with missing values
B. It models each feature with missing values as a function of other features, iterating until convergence
C. It only replaces with the mean
D. It requires complete data to begin

Answer: B

Q1392. Why can target encoding with naive implementation cause severe overfitting?

A. It removes too many features
B. Target information leaks into the encoded features, and rare categories get unreliable statistics that overfit to training noise
C. It always underfits
D. Target encoding is never problematic

Answer: B

Q1393. What is the impact of applying StandardScaler before train-test split versus after?

A. There is no difference either way
B. Applying before split causes data leakage because test set statistics influence the scaling parameters
C. Applying after split is always wrong
D. It only matters for deep learning

Answer: B

Q1394. How does ADASYN differ from SMOTE for handling class imbalance?

A. They are identical algorithms
B. ADASYN adaptively generates more synthetic samples for harder-to-learn minority instances near the decision boundary
C. ADASYN only works for regression
D. SMOTE generates more samples than ADASYN

Answer: B

Q1395. Why is preprocessing streaming data fundamentally different from batch preprocessing?

A. Streaming data is always cleaner
B. Streaming requires online or incremental algorithms that update statistics without access to the full dataset and must handle concept drift
C. There is no practical difference
D. Streaming data does not need preprocessing

Answer: B

Q1396. What is the optimal approach for preprocessing features with mixed distributions in a single dataset?

A. Apply the same transformation to all features
B. Use column transformers to apply different preprocessing pipelines to different feature subsets based on their characteristics
C. Always use StandardScaler for everything
D. Mixed distributions cannot be preprocessed

Answer: B

Q1397. How does quantile transformation make features follow a uniform or normal distribution?

A. It multiplies all values by a constant
B. It maps values to their cumulative distribution percentiles, then applies the inverse CDF of the target distribution
C. It sorts values alphabetically
D. It only works for binary features

Answer: B

Q1398. What challenges arise when preprocessing data with concept drift in production systems?

A. No challenges arise since preprocessing is static
B. Preprocessing statistics become stale as data distributions shift, requiring adaptive monitoring and periodic recalibration of transformers
C. Concept drift only affects model weights
D. Production systems never have concept drift

Answer: B

Q1399. Why is entity resolution important when integrating data from multiple sources?

A. It improves data encryption
B. It identifies and merges records referring to the same real-world entity despite different representations across sources
C. It is only relevant for social media data
D. Entity resolution deletes all duplicate data

Answer: B

Q1400. How does differential privacy noise injection during preprocessing protect sensitive data?

A. It makes data completely useless
B. It adds calibrated random noise that provides mathematical privacy guarantees while preserving aggregate statistical properties
C. It encrypts individual records
D. Noise injection always destroys data utility

Answer: B

Q1401. How can principal component analysis be used as an EDA tool for high-dimensional data?

A. PCA is only for model training
B. PCA reduces dimensions while preserving variance, enabling 2D/3D visualization that reveals clusters, outliers, and structure invisible in individual features
C. PCA removes all data variation
D. PCA only works for two features

Answer: B

Q1402. What is the Anscombe quartet and what lesson does it teach about EDA?

A. It is a type of neural network
B. Four datasets with identical summary statistics but drastically different distributions, demonstrating the necessity of visualization over summary statistics alone
C. It is a data compression algorithm
D. It is a type of hypothesis test

Answer: B

Q1403. How does mutual information provide a more general dependency measure than correlation in EDA?

A. Mutual information is identical to correlation
B. Mutual information captures any statistical dependency including nonlinear relationships, while correlation only measures linear association
C. It only works for binary variables
D. Mutual information is always zero

Answer: B

Q1404. Why might the Durbin-Watson statistic be crucial for EDA in regression analysis?

A. It measures data size
B. It detects autocorrelation in residuals, which violates regression assumptions and can lead to unreliable coefficient estimates and confidence intervals
C. It only applies to classification
D. It is a measure of accuracy

Answer: B

Q1405. What is the purpose of Cook's distance in regression EDA?

A. To measure cooking time for algorithms
B. To identify influential observations whose removal would substantially change the regression model's fitted values
C. To measure the distance between clusters
D. To compute feature importance

Answer: B

Q1406. How does the Kolmogorov-Smirnov test differ from the Anderson-Darling test for distribution testing?

A. They are identical tests
B. K-S measures the maximum difference between CDFs (sensitive to the center) while Anderson-Darling gives more weight to tails, better detecting tail deviations
C. K-S is always more powerful
D. Anderson-Darling cannot test normality

Answer: B

Q1407. What is the ecological fallacy and how can it mislead EDA conclusions?

A. It is a type of environmental data
B. Drawing conclusions about individuals based on aggregate group data, which can be misleading because group-level patterns may not hold for individuals
C. It only occurs in biological data
D. It is the same as Simpson's paradox

Answer: B

Q1408. How can SHAP summary plots enhance EDA by revealing feature-target relationships?

A. SHAP only works after model deployment
B. SHAP summary plots show each feature's impact on predictions across the dataset, revealing nonlinear relationships and interaction effects invisible in standard correlation analysis
C. SHAP replaces all other EDA methods
D. SHAP only works for linear models

Answer: B

Q1409. What is the Mahalanobis distance and why is it superior to Euclidean distance for multivariate outlier detection?

A. It is the same as Euclidean distance
B. It accounts for correlations and variances of features, correctly identifying outliers in correlated multivariate data where Euclidean distance fails
C. It only works for two dimensions
D. It ignores feature correlations

Answer: B

Q1410. How does the concept of data granularity affect EDA conclusions and subsequent modeling?

A. Granularity has no effect on analysis
B. The level of detail in data aggregation affects visible patterns, and analyzing at wrong granularity can hide important trends or create misleading artifacts
C. Finer granularity is always better
D. Granularity only matters for images

Answer: B

Q1411. What is the PAC learning framework and what does it guarantee?

A. It is a data format standard
B. It provides theoretical guarantees that a learning algorithm will probably approximately correctly learn a concept given sufficient data and computation
C. It is a type of neural network
D. PAC stands for Parallel Adaptive Computing

Answer: B

Q1412. How does Bayesian linear regression differ from ordinary least squares?

A. They are mathematically identical
B. Bayesian regression treats parameters as distributions rather than point estimates, incorporating prior beliefs and providing uncertainty quantification
C. Bayesian regression cannot handle multiple features
D. OLS always provides better predictions

Answer: B

Q1413. What is the dual formulation of SVM and why is it computationally advantageous?

A. It doubles the training time
B. The dual formulation expresses the problem in terms of inner products between data points, enabling the kernel trick and being more efficient when features outnumber samples
C. It is only for linear problems
D. The dual formulation always gives different results

Answer: B

Q1414. What is the relationship between logistic regression and maximum entropy models?

A. They are completely unrelated
B. Logistic regression is equivalent to a maximum entropy model with the same feature functions, both finding the least biased distribution consistent with observed data
C. Maximum entropy is only for physics
D. Logistic regression has higher entropy

Answer: B

Q1415. How does the Rademacher complexity bound relate to model generalization?

A. It measures training speed
B. It bounds the generalization gap by measuring how well the hypothesis class can fit random noise, with lower complexity indicating better generalization
C. It is a type of activation function
D. It only applies to unsupervised learning

Answer: B

Q1416. What is the difference between model stability and model accuracy in supervised learning?

A. They always improve together
B. Stability measures how much predictions change with small data perturbations while accuracy measures correctness, and there can be tradeoffs between them
C. They are identical metrics
D. Stability is irrelevant for good models

Answer: B

Q1417. Why can high-dimensional feature spaces make linear separability more likely?

A. They always make classification harder
B. Cover's theorem states that complex patterns are more likely to be linearly separable when projected into high-dimensional spaces
C. High dimensions reduce data quality
D. Linear separability is independent of dimensionality

Answer: B

Q1418. What is conformal prediction and how does it provide prediction guarantees?

A. It is a type of data augmentation
B. It produces prediction sets with a guaranteed coverage probability by using nonconformity scores from calibration data, requiring minimal distributional assumptions
C. It only works for regression
D. It is another name for Bayesian prediction

Answer: B

Q1419. How does quantile regression differ from ordinary least squares regression?

A. They produce identical outputs
B. Quantile regression estimates conditional quantiles rather than the conditional mean, providing a more complete picture of the response distribution
C. Quantile regression is simpler than OLS
D. OLS estimates all quantiles simultaneously

Answer: B

Q1420. What is the theory behind margin-based generalization bounds in classification?

A. Margins are irrelevant to generalization
B. Larger classification margins imply tighter generalization bounds because the model is more robust to perturbations, connecting SVM margins to PAC-Bayes theory
C. Smaller margins always generalize better
D. Margin theory only applies to linear models

Answer: B

Q1421. How does the second-order Taylor expansion in XGBoost's objective improve optimization?

A. It makes training slower
B. Using both first and second derivatives of the loss function provides curvature information for more accurate leaf value estimation and better splitting decisions
C. It eliminates the need for a loss function
D. Second-order expansion is identical to first-order

Answer: B

Q1422. What is the relationship between ensemble diversity and error reduction?

A. Diversity is irrelevant to error
B. The error of an ensemble is bounded by the average individual error minus a term proportional to diversity, so greater diversity leads to larger error reduction
C. Maximum diversity always gives zero error
D. Identical models give the best ensemble

Answer: B

Q1423. How does CatBoost's ordered boosting address target leakage in gradient boosting?

A. It does not address target leakage
B. It computes residuals using only previously seen observations in a random permutation, preventing the model from using future target information
C. It removes all categorical features
D. It uses the same ordering for all iterations

Answer: B

Q1424. What is the effect of correlation between base learners on ensemble variance reduction?

A. Correlation has no effect
B. Higher correlation between base learners reduces the variance reduction benefit, as the ensemble variance includes the average pairwise correlation term
C. High correlation improves ensembles
D. Correlation is impossible between ensemble members

Answer: B

Q1425. What is Shapley value-based feature importance and how does it improve upon impurity-based importance?

A. They are identical measures
B. Shapley values provide theoretically fair attribution by considering all possible feature coalitions, avoiding the bias toward high-cardinality features that affects impurity-based importance
C. Shapley values are faster to compute
D. Impurity-based importance is always more accurate

Answer: B

Q1426. Why can gradient boosting be interpreted as gradient descent in function space?

A. It cannot be interpreted that way
B. Each boosting iteration adds a function (tree) that approximates the negative gradient of the loss in function space, minimizing the loss by taking steps in the space of functions
C. It only uses numerical gradients
D. Function space interpretation is purely theoretical

Answer: B

Q1427. What is the computational complexity advantage of histogram-based gradient boosting?

A. It has no advantage
B. Binning continuous features into discrete histograms reduces split finding from O(n*features) to O(bins*features) per level, dramatically speeding up training on large datasets
C. It increases memory usage
D. It only works for categorical features

Answer: B

Q1428. How does the concept of complementarity differ from diversity in ensemble design?

A. They are identical concepts
B. Diversity means models make different errors, while complementarity means models' errors compensate each other in a way that benefits the combined prediction
C. Complementarity is about feature selection
D. Diversity is always sufficient for good ensembles

Answer: B

Q1429. What is the theoretical justification for why Random Forest's OOB error is a good estimate of generalization error?

A. OOB error is not a good estimate
B. Each sample is evaluated only by trees that did not include it in their bootstrap sample, making OOB predictions equivalent to cross-validation predictions
C. OOB always overestimates error
D. It is only valid for small datasets

Answer: B

Q1430. What is the multi-armed bandit approach to dynamic ensemble selection?

A. It is a type of card game
B. It treats each base model as an arm, dynamically selecting models for each prediction based on their recent performance using exploration-exploitation strategies
C. It always selects all models
D. It only works for binary classification

Answer: B

Q1431. What is the theoretical connection between K-Means and Gaussian Mixture Models?

A. They are completely unrelated
B. K-Means can be viewed as a special case of GMM where all covariances are equal isotropic matrices and cluster assignments become hard as variance approaches zero
C. GMM is a simpler version of K-Means
D. K-Means uses probabilistic assignments

Answer: B

Q1432. How does HDBSCAN improve upon DBSCAN for clusters of varying density?

A. It uses the same algorithm as DBSCAN
B. HDBSCAN builds a hierarchy of DBSCAN clusterings over all epsilon values and extracts the most stable clusters, automatically handling varying densities
C. It requires the user to specify more parameters
D. It only works for uniform density clusters

Answer: B

Q1433. Why does the curse of dimensionality particularly affect distance-based clustering algorithms?

A. It has no effect on clustering
B. In high dimensions, distances between points converge making clusters indistinguishable, and the volume of the space grows exponentially requiring exponentially more data
C. High dimensions always improve clustering
D. Distance-based algorithms are immune to dimensionality

Answer: B

Q1434. What is the information bottleneck method in unsupervised learning?

A. A physical bottleneck in data pipelines
B. A method that finds compressed representations that retain maximum mutual information with a relevant variable while discarding irrelevant information
C. It only applies to autoencoders
D. It is the same as PCA

Answer: B

Q1435. How does the Wasserstein distance provide a more meaningful comparison between distributions than KL divergence for clustering?

A. They are mathematically equivalent
B. Wasserstein distance accounts for the geometry of the space and is defined even when distributions have non-overlapping support, unlike KL divergence which becomes infinite
C. KL divergence is always better
D. Wasserstein distance cannot be computed

Answer: B

Q1436. What is the relationship between spectral clustering and graph partitioning?

A. They are unrelated
B. Spectral clustering constructs a similarity graph, then uses eigenvalues of the graph Laplacian to partition it, making clustering equivalent to the normalized graph cut problem
C. Spectral clustering does not use graphs
D. Graph partitioning is a supervised method

Answer: B

Q1437. Why is the Evidence Lower Bound (ELBO) used as the training objective in variational autoencoders?

A. It is simpler than reconstruction error
B. The true log-likelihood is intractable, so ELBO provides a tractable lower bound that combines reconstruction quality with regularization of the latent space
C. ELBO is the same as MSE loss
D. It is only used for classification

Answer: B

Q1438. What is the stability-based approach to determining the number of clusters?

A. It always selects K=2
B. It evaluates how consistently clustering solutions are reproduced under data perturbations, with the correct K yielding the most stable results
C. It measures computational stability
D. It only works for DBSCAN

Answer: B

Q1439. How does deep embedded clustering combine deep learning with clustering?

A. It uses a single K-Means step
B. It jointly optimizes a deep autoencoder's feature representation and cluster assignments, iteratively refining both the embedding and clusters
C. It only uses linear transformations
D. It is identical to standard K-Means

Answer: B

Q1440. What is the fundamental problem of identifiability in mixture models?

A. Mixture models always produce unique solutions
B. The likelihood function has multiple equivalent optima due to label switching and potential parameter degeneracies, making the solution non-unique
C. Identifiability only matters for deep learning
D. Mixture models have no parameters

Answer: B

Q1441. What is the relationship between the Brier score and model calibration?

A. Brier score only measures discrimination
B. The Brier score can be decomposed into calibration, refinement, and uncertainty components, making it sensitive to both calibration and discrimination
C. Brier score ignores calibration
D. It is only useful for multi-class problems

Answer: B

Q1442. Why might the McNemar test be more appropriate than paired t-test for comparing classifiers?

A. They test the same thing
B. McNemar's test compares classifiers based on their disagreement patterns on specific samples rather than aggregate metrics, providing more statistical power
C. McNemar test requires more data
D. Paired t-test is always more powerful

Answer: B

Q1443. What is the purpose of permutation testing for model evaluation?

A. To permute the features
B. To establish a null distribution by repeatedly shuffling labels and retraining, determining if the model's performance is significantly better than chance
C. It always validates the model
D. Permutation tests only work for linear models

Answer: B

Q1444. How does the concept of distribution shift affect model evaluation in production?

A. Distribution shift never occurs in practice
B. When production data differs from training/test data distribution, held-out test performance overestimates real-world performance, requiring ongoing monitoring
C. Distribution shift only affects accuracy
D. It can be completely prevented by larger test sets

Answer: B

Q1445. What is the purpose of isotonic regression in model calibration?

A. It is a type of linear regression
B. Isotonic regression fits a non-decreasing step function to map model scores to calibrated probabilities, more flexible than Platt scaling
C. It always produces worse calibration
D. It is only used for regression models

Answer: B

Q1446. What is the effective sample size problem in cross-validation and how does it affect reliability?

A. Cross-validation always gives reliable estimates
B. When data is limited, cross-validation folds share training data making fold estimates correlated, violating independence assumptions and causing underestimation of variance
C. It only affects large datasets
D. Effective sample size is always equal to dataset size

Answer: B

Q1447. Why is the area under the precision-recall curve more informative than ROC-AUC for highly imbalanced problems?

A. ROC-AUC is always more informative
B. PR-AUC is sensitive to the absolute number of false positives which dominates in imbalanced settings, while ROC-AUC can be misleadingly high due to the large number of true negatives
C. PR-AUC and ROC-AUC always agree
D. PR-AUC ignores the positive class

Answer: B

Q1448. What is the difference between prospective and retrospective evaluation of ML models?

A. They are the same approach
B. Prospective evaluation tests on future data collected after model training while retrospective uses historical held-out data, with prospective being more reliable for deployment
C. Retrospective is always better
D. Prospective evaluation uses training data

Answer: B

Q1449. How does the concept of metric elicitation help select the right evaluation metric?

A. It always selects accuracy
B. Metric elicitation formally derives the optimal evaluation metric from stakeholder preferences about the relative costs of different types of errors
C. It selects metrics randomly
D. It only applies to unsupervised learning

Answer: B

Q1450. What is the limitation of using a single train-test split for model evaluation and how does repeated stratified K-fold address it?

A. Single splits are always sufficient
B. A single split gives high variance estimates dependent on the specific random partition, while repeated stratified K-fold averages over multiple splits and preserves class ratios for more stable estimates
C. Repeated K-fold is always worse
D. Single splits are only bad for small datasets

Answer: B

Q1451. How does the Boruta algorithm determine feature relevance compared to regular feature importance?

A. It uses the same approach as Random Forest importance
B. Boruta creates shadow features by shuffling original features and compares each real feature's importance against the maximum shadow importance over multiple iterations
C. It only removes constant features
D. It is faster than simple importance ranking

Answer: B

Q1452. What is the concept of feature store consistency between training and serving?

A. It means using the same programming language
B. Feature store consistency ensures the same feature computation logic and transformations are applied during both model training and real-time inference, preventing training-serving skew
C. It only matters for batch processing
D. Consistency is guaranteed automatically

Answer: B

Q1453. Why might learned embeddings outperform manually engineered features for categorical variables?

A. Manual features are always superior
B. Embeddings learn dense representations that capture latent semantic relationships between categories in the context of the specific task, beyond what manual encoding schemes can express
C. Embeddings only work for text
D. Manual features capture more relationships

Answer: B

Q1454. How does automated feature engineering with deep feature synthesis work?

A. It uses deep learning only
B. DFS automatically generates features by stacking aggregation and transformation primitives across relational tables, systematically exploring the feature space
C. It only creates simple averages
D. It requires no data

Answer: B

Q1455. What is the theoretical basis for why polynomial feature expansion can lead to overfitting?

A. Polynomials never cause overfitting
B. The VC dimension grows rapidly with polynomial degree, increasing model capacity exponentially and requiring exponentially more data to avoid overfitting
C. Polynomials always underfit
D. Overfitting is unrelated to feature expansion

Answer: B

Q1456. How does entity embedding of categorical variables from neural networks transfer to tree-based models?

A. It cannot be transferred
B. Entity embeddings learned by neural networks can be extracted and used as numerical features in tree-based models, combining the representation learning of deep learning with the strengths of gradient boosting
C. Tree models always outperform embeddings
D. Neural embeddings are incompatible with trees

Answer: B

Q1457. What is the stability selection method and how does it improve feature selection reliability?

A. It selects the most stable features across time
B. Stability selection repeatedly subsamples data and applies feature selection, keeping features that are consistently selected across many subsamples, controlling false discovery rate
C. It uses the same selection every time
D. It only works for regression

Answer: B

Q1458. Why is the order of feature engineering steps important in a preprocessing pipeline?

A. Order never matters
B. Transformations are not commutative: scaling before encoding produces different results than encoding before scaling, and incorrect ordering can introduce data leakage or distort distributions
C. All orderings produce identical results
D. Only the last step matters

Answer: B

Q1459. How do feature interactions affect the performance gap between linear and nonlinear models?

A. Interactions are irrelevant to model choice
B. When important feature interactions exist, linear models require explicit interaction features while tree-based models discover them automatically, explaining their performance advantage
C. Linear models automatically detect interactions
D. Nonlinear models cannot use interactions

Answer: B

Q1460. What is the concept of leakage-free feature engineering in production ML systems?

A. It is impossible to prevent leakage
B. Ensuring all features are computed using only information available at prediction time and that no future or target information contaminates features during training or serving
C. Leakage only occurs in training
D. Production systems never have leakage

Answer: B

Q1461. What is the connection between information theory and deep learning through the Information Bottleneck principle?

A. Information theory is irrelevant to deep learning
B. Each layer progressively compresses input information while retaining information relevant to the output, balancing compression and prediction
C. Information always increases through layers
D. The Information Bottleneck only applies to autoencoders

Answer: B

Q1462. What is the double descent phenomenon in deep learning?

A. Models always improve with more parameters
B. Test error shows a classical U-shaped curve then decreases again in the over-parameterized regime, challenging classical bias-variance intuitions
C. It only occurs with small datasets
D. Double descent means training the model twice

Answer: B

Q1463. How do implicit biases of gradient descent favor certain solutions in over-parameterized networks?

A. Gradient descent always finds the worst solution
B. SGD implicitly regularizes toward flat minima with small norm weights, which tend to generalize better, even without explicit regularization
C. Gradient descent has no bias
D. Implicit bias only affects small networks

Answer: B

Q1464. What is the Sharpness-Aware Minimization technique and why does it improve generalization?

A. It sharpens image inputs
B. SAM seeks parameters in regions where the loss landscape is uniformly low rather than just at a single low point, finding flatter minima that generalize better
C. It makes the loss function sharper
D. SAM only works for classification

Answer: B

Q1465. What is the role of the Hessian spectrum in understanding neural network optimization?

A. The Hessian is not relevant to optimization
B. The eigenvalue distribution of the Hessian reveals the loss landscape geometry, with a few large eigenvalues and many near-zero values indicating low effective dimensionality
C. The Hessian spectrum is always flat
D. It only applies to linear models

Answer: B

Q1466. How does neural network pruning maintain accuracy while reducing model size?

A. Pruning always degrades accuracy significantly
B. Over-parameterized networks contain redundant connections, and structured or unstructured pruning can remove a large fraction of weights with minimal accuracy loss, especially with fine-tuning
C. Pruning only removes input features
D. It only works for convolutional networks

Answer: B

Q1467. What is the difference between post-training quantization and quantization-aware training?

A. They are identical approaches
B. Post-training quantization converts a trained model's weights to lower precision, while quantization-aware training simulates quantization during training for better accuracy retention
C. Post-training is always better
D. Quantization-aware training uses higher precision

Answer: B

Q1468. Why does batch normalization enable higher learning rates?

A. It has no effect on learning rates
B. By normalizing layer inputs, batch norm reduces internal covariate shift and smooths the loss landscape, allowing larger learning rate steps without divergence
C. It only affects the first layer
D. Batch norm always reduces learning rates

Answer: B

Q1469. What is the grokking phenomenon in deep learning?

A. A standard training pattern
B. A delayed generalization where a model memorizes training data first and only much later suddenly learns to generalize, well past the point of overfitting
C. Grokking means fast convergence
D. It only occurs in shallow networks

Answer: B

Q1470. How does the loss of plasticity problem affect continually trained neural networks?

A. Plasticity always increases with training
B. Networks gradually lose the ability to learn new tasks as training progresses, due to dead neurons, saturated weights, and reduced effective learning capacity
C. Loss of plasticity only affects small models
D. It is the same as catastrophic forgetting

Answer: B

Q1471. How does the Transformer's multi-head attention improve upon single-head attention?

A. It uses more memory without benefit
B. Multiple heads attend to different representation subspaces at different positions simultaneously, capturing diverse relational patterns that a single head cannot
C. More heads always reduce performance
D. Multi-head is computationally cheaper

Answer: B

Q1472. What is the theoretical basis behind why deeper networks can represent functions more efficiently than wider shallow ones?

A. Depth has no theoretical advantage
B. Depth enables compositional representation where complex functions are built hierarchically from simpler sub-functions, achieving exponential efficiency over shallow networks for certain function classes
C. Wider networks are always more efficient
D. Deep and shallow networks have equal capacity

Answer: B

Q1473. How does the score-based diffusion model formulation connect to stochastic differential equations?

A. They are unrelated
B. Diffusion models can be formulated as continuous-time SDEs where the forward process adds noise and the reverse process removes it, with the score function learned by the neural network guiding denoising
C. SDEs are only for financial modeling
D. Diffusion models use deterministic equations only

Answer: B

Q1474. What is the fundamental trade-off in designing efficient Transformer architectures?

A. There is no trade-off
B. Reducing the quadratic O(n^2) attention complexity to linear or sub-quadratic while preserving the model's ability to capture long-range dependencies and global context
C. Efficiency always improves accuracy
D. Transformer efficiency only depends on model size

Answer: B

Q1475. How does LoRA enable efficient fine-tuning of large pre-trained models?

A. It retrains all parameters
B. LoRA freezes pre-trained weights and injects trainable low-rank decomposition matrices into attention layers, dramatically reducing trainable parameters while maintaining performance
C. It deletes unused parameters
D. LoRA only works for small models

Answer: B

Q1476. What is the role of the exponential moving average of model weights in improving generalization?

A. It speeds up training only
B. EMA maintains a smoothed version of weights that averages over the training trajectory, effectively ensembling models from different training stages and finding flatter minima
C. EMA always reduces accuracy
D. It is only used during data preprocessing

Answer: B

Q1477. Why do Vision Transformers partition images into patches rather than processing individual pixels?

A. Patches are visually more appealing
B. Processing individual pixels would make the self-attention sequence length equal to image resolution squared, making computation prohibitively expensive, while patches create manageable sequences
C. Patches always produce better accuracy
D. Pixel-level processing is computationally cheaper

Answer: B

Q1478. How does the Mixture of Experts architecture achieve conditional computation?

A. All experts process every input
B. A gating network routes each input to only a small subset of expert networks, so total computation scales sub-linearly with model size
C. Experts are selected randomly without any gating
D. MoE reduces model capacity

Answer: B

Q1479. What is the relationship between contrastive learning and the InfoNCE loss?

A. They are unrelated
B. InfoNCE is the standard loss function for contrastive learning that maximizes a lower bound on mutual information between positive pairs while treating other samples as negatives
C. InfoNCE is only for supervised learning
D. Contrastive learning does not use loss functions

Answer: B

Q1480. How does the concept of neural scaling laws inform model and dataset design decisions?

A. Scaling laws are not useful for planning
B. Empirical power-law relationships between model size, dataset size, compute budget, and loss enable predicting performance of larger models from smaller experiments, guiding resource allocation
C. Performance scales linearly with compute
D. Scaling laws only apply to language models

Answer: B

Q1481. What is the fundamental challenge of evaluating open-ended text generation with automatic metrics?

A. Automatic metrics are always perfect
B. There are many valid ways to express the same meaning, so reference-based metrics like BLEU penalize valid but differently-worded outputs, failing to capture semantic equivalence
C. All generated text can be evaluated by word matching
D. Open-ended generation does not need evaluation

Answer: B

Q1482. How does the rotary position embedding encode positional information in Transformers?

A. It adds a fixed vector to all positions
B. RoPE rotates the query and key vectors by position-dependent angles, encoding relative positions through the dot product's natural decay with distance
C. It uses learned position embeddings
D. RoPE removes all position information

Answer: B

Q1483. What is the constitutional AI approach to aligning language models?

A. Writing an AI constitution document
B. Training the model to critique and revise its own outputs according to a set of principles, reducing reliance on human feedback for each specific case
C. It is a type of government regulation
D. Constitutional AI replaces all human oversight

Answer: B

Q1484. What is the difference between sparse and dense retrieval in retrieval-augmented generation?

A. They produce identical results
B. Sparse retrieval uses term matching like BM25 while dense retrieval uses learned semantic embeddings, with dense capturing meaning beyond exact word overlap
C. Sparse retrieval is always better
D. Dense retrieval only matches exact words

Answer: B

Q1485. How does Direct Preference Optimization simplify the RLHF pipeline?

A. It adds more complexity than RLHF
B. DPO directly optimizes the language model policy using preference pairs without needing a separate reward model or reinforcement learning, simplifying the training pipeline
C. DPO requires a separate reward model
D. It is identical to standard RLHF

Answer: B

Q1486. What is the mixture of softmaxes approach and how does it address the softmax bottleneck?

A. It mixes multiple models together
B. It uses a mixture of multiple softmax distributions to increase the rank of the output distribution matrix, addressing the limitation that a single softmax cannot model complex distributions
C. Single softmax has no limitations
D. It reduces model capacity

Answer: B

Q1487. Why is tokenizer choice critically important for multilingual NLP models?

A. Tokenizers are identical across languages
B. The tokenizer determines how efficiently each language is represented: languages with poor tokenizer coverage get fragmented into more tokens, increasing compute cost and reducing effective context length
C. Tokenizers only affect English text
D. All languages require the same number of tokens

Answer: B

Q1488. What is the in-context learning phenomenon in large language models?

A. Learning that occurs during pre-training
B. LLMs can learn to perform new tasks from examples provided in the prompt without any parameter updates, implicitly implementing learning algorithms within their forward pass
C. It requires fine-tuning the model
D. In-context learning only works for classification

Answer: B

Q1489. How does the KV-cache optimization speed up autoregressive Transformer inference?

A. It caches the entire model on disk
B. It stores previously computed key and value tensors to avoid redundant recomputation at each generation step, reducing cumulative attention computation
C. KV-cache only works during training
D. It caches output probabilities

Answer: B

Q1490. What is the challenge of faithfulness in abstractive summarization systems?

A. Faithfulness is not important for summaries
B. Abstractive models can generate fluent but factually incorrect summaries that contradict or fabricate information not present in the source document
C. Extractive summarization has worse faithfulness
D. All neural summaries are perfectly faithful

Answer: B

Q1491. How do deformable convolutions improve upon standard convolutions for object detection?

A. They use fixed rectangular receptive fields
B. Deformable convolutions learn input-dependent offsets for sampling positions, enabling adaptive receptive fields that conform to object shapes
C. They always produce worse results
D. Deformable convolutions are faster than standard convolutions

Answer: B

Q1492. What is the concept of equivariance versus invariance in CNN design?

A. They are the same property
B. Equivariance means the output transforms predictably with input transformations while invariance means the output remains unchanged
C. CNNs are always invariant to all transformations
D. Equivariance prevents learning

Answer: B

Q1493. Why is the anchor-free approach gaining popularity over anchor-based methods in object detection?

A. Anchors are always better
B. Anchor-free methods directly predict object centers and sizes without predefined anchor boxes, reducing hyperparameters and simplifying the pipeline
C. Anchor-free methods cannot detect objects
D. All modern detectors use anchors

Answer: B

Q1494. What is the fundamental challenge of training object detection models with highly imbalanced foreground-background ratios?

A. Imbalance does not affect detection
B. The vast majority of anchor locations are background, which dominate the loss and overwhelm the gradient signal from sparse foreground objects
C. All locations are equally important
D. Background samples improve foreground detection

Answer: B

Q1495. How does the CLIP model enable zero-shot image classification?

A. It memorizes all possible images
B. CLIP learns a shared embedding space for images and text through contrastive pre-training, enabling classification by comparing image embeddings with text descriptions of classes
C. Zero-shot classification is impossible
D. CLIP only works for text

Answer: B

Q1496. What is the role of attention mechanisms in modern image segmentation architectures?

A. Attention is not used in segmentation
B. Attention modules capture long-range dependencies between distant pixels, enabling global context aggregation critical for understanding semantic relationships
C. Attention only works for small images
D. It replaces all convolutional layers

Answer: B

Q1497. Why is test-time augmentation effective for improving computer vision model predictions?

A. It has no effect on predictions
B. TTA applies multiple augmentations to the test image, runs inference on each, and aggregates predictions, effectively ensembling predictions from different perspectives
C. It only works during training
D. TTA always reduces accuracy

Answer: B

Q1498. How do neural radiance fields represent 3D scenes from 2D images?

A. They stack 2D images together
B. NeRF trains a neural network to map 3D coordinates and viewing direction to color and density, then uses volume rendering to synthesize novel views
C. They require 3D scanners
D. NeRF only works for simple objects

Answer: B

Q1499. What is the challenge of domain gap in transferring CV models across different visual domains?

A. Domain gap does not exist
B. Models trained on one visual domain perform poorly on another due to differences in appearance, lighting, style, and sensor characteristics, requiring domain adaptation techniques
C. All visual domains look identical to CNNs
D. Domain gap only affects text models

Answer: B

Q1500. What are the tradeoffs in the CAP theorem and how do modern distributed databases address them?

A. There are no tradeoffs in distributed systems
B. The CAP theorem states only two of consistency, availability, and partition tolerance can be fully guaranteed. Modern systems like CockroachDB and Spanner make nuanced tradeoffs using techniques like consensus protocols
C. All three guarantees are always achievable
D. CAP theorem is no longer relevant

Answer: B

Q1501. How does Apache Flink achieve exactly-once processing semantics in stream processing?

A. It simply processes each message once
B. Flink uses distributed snapshots (Chandy-Lamport algorithm) for consistent checkpointing combined with two-phase commit for external sinks to guarantee exactly-once end-to-end processing
C. Exactly-once is impossible in streaming
D. It buffers all data in memory

Answer: B

Q1502. What is the challenge of data skew in distributed joins and how can it be mitigated?

A. Data skew has no impact on joins
B. Skewed key distributions cause some partitions to receive disproportionately more data, creating bottleneck tasks. Mitigations include salted keys, broadcast joins for small tables, and adaptive query engines
C. All data is always evenly distributed
D. Skew only affects small datasets

Answer: B

Q1503. How does the Delta Lake transaction log provide ACID guarantees on top of object storage?

A. It uses a traditional relational database
B. Delta Lake maintains a transaction log of ordered JSON files recording every change, using optimistic concurrency control to provide atomicity, consistency, isolation, and durability on cloud object stores
C. ACID is impossible on object storage
D. It modifies the underlying storage system

Answer: B

Q1504. What is the theoretical basis for why bloom filters are space-efficient for approximate membership testing?

A. Bloom filters store exact data
B. Bloom filters use multiple hash functions mapping elements to a bit array, allowing false positives but no false negatives with space usage sub-linear in the number of elements
C. They always produce exact results
D. Bloom filters require more space than hash tables

Answer: B

Q1505. Why is backpressure management critical in streaming data systems?

A. Backpressure is not a concern in streaming
B. When downstream consumers cannot keep up with upstream producers, backpressure prevents system overload by propagating flow control signals to slow producers or buffer data appropriately
C. Streaming systems never have speed mismatches
D. Backpressure always causes data loss

Answer: B

Q1506. How does the concept of materialized views improve query performance in analytical databases?

A. They have no performance benefit
B. Materialized views pre-compute and store query results, trading storage space for query speed by avoiding re-computation of expensive aggregations and joins
C. They only work for simple queries
D. Materialized views always have stale data

Answer: B

Q1507. What is the log-structured merge tree and why is it used in write-heavy big data systems?

A. It is a type of logging framework
B. LSM trees buffer writes in memory, flush sorted segments to disk, and merge them periodically, optimizing write throughput at the cost of read amplification
C. It only works for read-heavy workloads
D. LSM trees are slower than B-trees for all operations

Answer: B

Q1508. How does federated query processing enable analytics across heterogeneous data sources?

A. It requires all data to be in one database
B. Federated query engines translate a single query into sub-queries for different data sources, execute them in parallel, and combine results, enabling analytics without data migration
C. It only works with SQL databases
D. Federated queries are always slower than centralized

Answer: B

Q1509. What is the challenge of maintaining data quality at scale in big data systems?

A. Data quality is automatic at scale
B. Scale amplifies data quality issues: schema drift across sources, late-arriving data, duplicate events, and silent failures require automated validation, schema registries, dead letter queues, and data observability platforms
C. Big data is always high quality
D. Data quality only matters for small datasets

Answer: B

Q1510. What is the training-serving skew problem and how does a feature store address it?

A. There is no such problem
B. Training-serving skew occurs when feature computation differs between training and serving, causing silent performance degradation. Feature stores provide a single source of truth for feature definitions used in both contexts
C. Feature stores cause more skew
D. Skew only affects small models

Answer: B

Q1511. How does continuous training differ from traditional periodic retraining in production ML?

A. They are identical approaches
B. Continuous training automatically triggers model retraining based on data changes, performance degradation, or schedules, with automated validation and deployment, creating a self-maintaining ML system
C. Continuous training never stops training
D. Traditional retraining is always better

Answer: B

Q1512. What are the unique challenges of testing ML systems compared to traditional software?

A. ML systems are easier to test
B. ML systems have non-deterministic outputs, depend on training data quality, require testing of data pipelines and model behavior, and lack clear oracles for correct outputs in many cases
C. Testing is identical for ML and software
D. ML systems do not need testing

Answer: B

Q1513. How does the concept of ML technical debt manifest in production systems?

A. ML systems have no technical debt
B. ML systems accumulate hidden debt through entangled features, undeclared data dependencies, feedback loops, configuration complexity, and the difficulty of monitoring and testing non-deterministic systems
C. Technical debt is only a software concern
D. ML debt is always visible

Answer: B

Q1514. What is the role of a serving graph in complex ML deployment architectures?

A. A graph showing model accuracy
B. A serving graph orchestrates multiple models and processing steps into a single prediction pipeline, handling feature transformation, model ensemble, post-processing, and fallback logic
C. It only serves single models
D. Serving graphs are only theoretical

Answer: B

Q1515. How does data-centric AI differ from the traditional model-centric approach to improving ML systems?

A. They are the same approach
B. Data-centric AI focuses on systematically improving the quality and quantity of training data rather than iterating on model architectures, recognizing data quality as the primary performance bottleneck
C. Data-centric ignores data quality
D. Model-centric is always superior

Answer: B

Q1516. What is the challenge of implementing real-time feature computation for online model serving?

A. Real-time features are simple to implement
B. Real-time features require low-latency computation from streaming data, consistent logic with training features, handling of late-arriving data, and maintaining state across events
C. Real-time features are identical to batch features
D. Latency does not matter for serving

Answer: B

Q1517. How does ML governance address compliance and regulatory requirements for AI systems?

A. Governance is unnecessary for ML
B. ML governance establishes policies, processes, and tools for model documentation, approval workflows, bias auditing, access control, and audit trails to meet regulatory requirements
C. Governance only applies to financial models
D. All ML models are automatically compliant

Answer: B

Q1518. What is the multi-armed bandit approach to model selection in production?

A. A gambling strategy
B. Multi-armed bandit algorithms dynamically allocate traffic between model variants based on their observed performance, balancing exploration and exploitation more efficiently than static A/B tests
C. It always selects the same model
D. It requires manual selection

Answer: B

Q1519. Why is observability more important than traditional monitoring for production ML systems?

A. Observability and monitoring are identical
B. Observability enables investigation of novel, unexpected failure modes using high-cardinality telemetry data, while monitoring only checks predefined metrics against thresholds and cannot diagnose unknown unknowns
C. Monitoring covers all failure modes
D. Observability is only for infrastructure

Answer: B

Q1520. Why is the impossibility theorem for fairness significant for AI system design?

A. The theorem has been disproven
B. The impossibility theorem shows that several intuitive fairness criteria cannot all be satisfied simultaneously, requiring explicit tradeoff decisions in AI system design
C. All fairness criteria can be satisfied
D. The theorem only applies to binary classification

Answer: B

Q1521. How does machine unlearning attempt to address the right to be forgotten in trained AI models?

A. It deletes the model entirely
B. Machine unlearning develops algorithms to efficiently remove the influence of specific training data points from a trained model without full retraining, satisfying data deletion requests
C. Unlearning is impossible for neural networks
D. Simply deleting data files is sufficient

Answer: B

Q1522. What is the concept of counterfactual fairness in AI decision-making?

A. Making decisions based on fictional scenarios
B. A decision is counterfactually fair if it would remain the same in a hypothetical world where the individual's protected attribute had been different, ensuring the attribute had no causal influence
C. It means treating all groups identically
D. Counterfactual fairness ignores protected attributes

Answer: B

Q1523. How does the concept of differential privacy provide mathematical guarantees for data protection?

A. It encrypts all data permanently
B. Differential privacy ensures that the output of an analysis is statistically indistinguishable whether or not any single individual's data is included, with the privacy budget epsilon quantifying the guarantee
C. Differential privacy eliminates all utility
D. It only works for numerical data

Answer: B

Q1524. What is the challenge of ensuring AI transparency across the full ML pipeline?

A. Transparency is only needed for the model itself
B. Full transparency requires documenting data collection and consent, preprocessing decisions, model selection rationale, training procedures, evaluation methodology, deployment context, and ongoing monitoring, which is complex and costly
C. Transparency means making models open source
D. Full transparency is easily achieved

Answer: B

Q1525. How do adversarial attacks on AI systems raise unique ethical concerns?

A. Adversarial attacks have no ethical implications
B. Adversarial vulnerabilities mean AI systems can be deliberately manipulated to make dangerous errors, raising concerns about safety-critical deployment, security, and the responsibility gap when AI is deceived
C. Adversarial attacks only affect image models
D. All AI systems are robust to adversarial attacks

Answer: B

Q1526. What is the tension between model performance optimization and fairness in high-stakes AI applications?

A. There is no tension between performance and fairness
B. Optimizing for overall accuracy may come at the cost of disparate performance across demographic groups, requiring explicit fairness constraints that may reduce overall accuracy but improve equity
C. Fairness always improves accuracy
D. Performance and fairness are identical objectives

Answer: B

Q1527. How does the EU AI Act's risk-based framework categorize AI systems?

A. All AI systems are treated equally
B. The Act categorizes AI into unacceptable risk (banned), high risk (strict requirements including conformity assessment), limited risk (transparency obligations), and minimal risk (no specific obligations)
C. It only regulates social media AI
D. The Act has no enforcement mechanism

Answer: B

Q1528. What is the concept of AI sovereignty and why is it geopolitically significant?

A. AI that is self-aware
B. National control over AI infrastructure, data, research capabilities, and regulatory frameworks, which has become a strategic priority affecting global power dynamics and technology independence
C. It only matters for military AI
D. AI sovereignty is not a real concept

Answer: B

Q1529. How does the concept of meaningful human control apply to autonomous AI systems?

A. Humans must manually approve every AI decision
B. Meaningful human control requires that humans can understand, intervene in, and override AI decisions at appropriate levels while maintaining effective oversight even as AI systems become more autonomous
C. Human control is unnecessary for advanced AI
D. It means AI should never make decisions

Answer: B

Q1530. How does the Segment Anything Model generalize to unseen segmentation tasks?

A. It memorizes all possible object shapes
B. SAM is trained on a massive diverse dataset with a promptable architecture that accepts points, boxes, or text to segment any object without task-specific fine-tuning
C. It uses a fixed set of object categories
D. It only works for previously seen objects

Answer: B