

Question 48: In Data Science, what is the process of reducing the number of dimensions in a dataset called?

- A) Dimensionality reduction
- B) Dimension scaling
- C) Dimension transformation
- D) Dimension augmentation

Answer: A) Dimensionality reduction

Question 49: Which Python library is commonly used for data visualization?

- A) NumPy
- B) Pandas
- C) Seaborn
- D) Scikit-learn

Answer: C) Seaborn

Question 50: What is the process of dividing a dataset into three parts: training, validation, and testing sets, called?

- A) Data splitting
- B) Data preprocessing
- C) Data augmentation
- D) Data scaling

Answer: A) Data splitting

Question 51: Which statistical test is used to determine if there is a significant relationship between two categorical variables?

- A) Chi-square test
- B) T-test
- C) ANOVA (Analysis of Variance)
- D) Pearson correlation coefficient

Answer: A) Chi-square test

Question 52: What is the primary purpose of one-hot encoding in Data Science?

- A) To handle missing data
- B) To convert categorical data into numerical format
- C) To normalize the data

D) To standardize the data

Answer: B) To convert categorical data into numerical format

Question 53: In machine learning, what is the process of dividing a dataset into multiple subsets to improve the model's generalization called?

A) Cross-validation

B) Overfitting

C) Underfitting

D) Regularization

Answer: A) Cross-validation

Question 54: Which machine learning algorithm is used for anomaly detection in time series data?

A) Linear Regression

B) K-Means

C) Isolation Forest

D) Decision Trees

Answer: C) Isolation Forest

Question 55: What is the process of transforming data into a standardized scale called?

A) Data encoding

B) Data scaling

C) Data binning

D) Data imputation

Answer: B) Data scaling

Question 56: Which Python library is commonly used for numerical computing and scientific computing?

A) TensorFlow

B) PyTorch

C) NumPy

D) Scikit-learn

Answer: C) NumPy

Question 57: In Data Science, what is the process of combining multiple datasets into one called?

A) Data transformation

B) Data integration

C) Data preprocessing

D) Data augmentation

Answer: B) Data integration

Question 58: What is the process of dividing a dataset into homogeneous subgroups called?

A) Clustering

B) Classification

C) Regression

D) Dimensionality reduction

Answer: A) Clustering

Question 59: In machine learning, what is the process of assigning weights to different features to make them equally important called?

A) Feature engineering

B) Feature selection

C) Feature scaling

D) Feature weighting

Answer: D) Feature weighting

Question 60: What is the primary purpose of the sigmoid function in logistic regression?

A) To compute the loss function

B) To reduce overfitting

C) To compute the gradient descent

D) To produce probabilities for binary classification

Answer: D) To produce probabilities for binary classification

Question 61: Which machine learning algorithm is commonly used for regression problems with a large number of features?

A) Ridge Regression

B) Logistic Regression

C) Support Vector Machines (SVM)

D) K-Nearest Neighbors (KNN)

Answer: A) Ridge Regression

Question 62: In Data Science, what is the process of transforming numerical variables into a similar scale called?

- A) Data encoding
- B) Data scaling
- C) Data binning
- D) Data imputation

Answer: B) Data scaling

Question 63: Which Python library is commonly used for natural language processing (NLP) tasks like sentiment analysis and text classification?

- A) Pandas
- B) NumPy
- C) Seaborn
- D) NLTK (Natural Language Toolkit)

Answer: D) NLTK (Natural Language Toolkit)

Question 64: What is the process of transforming data into a format suitable for modeling called?

- A) Data preprocessing
- B) Data wrangling
- C) Data augmentation
- D) Data transformation

Answer: A) Data preprocessing

Question 65: In machine learning, what is the process of combining multiple weak learners to create a more powerful model called?

- A) Ensemble learning
- B) Transfer learning
- C) Reinforcement learning
- D) Supervised learning

Answer: A) Ensemble learning

Question 66: Which statistical test is used to determine if there is a significant relationship between two continuous variables?

- A) Chi-square test
- B) T-test
- C) Pearson correlation coefficient
- D) ANOVA (Analysis of Variance)

Answer: C) Pearson correlation coefficient

Question 67: What is the primary goal of k-fold cross-validation in machine learning?

- A) To avoid overfitting
- B) To handle missing data
- C) To reduce model complexity
- D) To improve feature selection

Answer: A) To avoid overfitting

Question 68: Which machine learning algorithm is used for reducing the dimensionality of high-dimensional data?

- A) Decision Trees
- B) K-Means
- C) Principal Component Analysis (PCA)
- D) Naive Bayes

Answer: C) Principal Component Analysis (PCA)

Question 69: In Data Science, what is the process of replacing outliers with more reasonable values called?

- A) Data imputation
- B) Data cleaning
- C) Data normalization
- D) Data transformation

Answer: A) Data imputation

Question 70: Which Python library is commonly used for deep learning and building artificial neural networks?

- A) TensorFlow
- B) Scikit-learn
- C) PyTorch
- D) NLTK (Natural Language Toolkit)

Answer: A) TensorFlow

Question 71: What is the primary purpose of the Lasso regression technique in machine learning?

- A) To improve model interpretability
- B) To handle missing data

C) To reduce the complexity of the model

D) To increase the accuracy of the model

Answer: C) To reduce the complexity of the model

Question 72: In Data Science, what is the process of transforming data into a normalized range called?

A) Data encoding

B) Data scaling

C) Data binning

D) Data augmentation

Answer: B) Data scaling

Question 73: Which machine learning algorithm is commonly used for clustering?

A) Decision Trees

B) Linear Regression

C) K-Means

D) Logistic Regression

Answer: C) K-Means

Question 74: What is the process of converting text data into numerical representations by counting the frequency of words called?

A) Tokenization

B) Bag-of-Words

C) Lemmatization

D) Stemming

Answer: B) Bag-of-Words

Question 75: In machine learning, what is the process of dividing a dataset into two parts: training and testing sets, called?

A) Data preprocessing

B) Data augmentation

C) Data splitting

D) Data transformation

Answer: C) Data splitting

Question 76: Which statistical measure is used to measure the strength and direction of a linear relationship between two variables?

- A) Mean
- B) Median
- C) Variance
- D) Correlation coefficient

Answer: D) Correlation coefficient

Question 77: What is the primary goal of the Kullback-Leibler (KL) divergence in machine learning?

- A) To measure the distance between two probability distributions
- B) To compute the gradient descent
- C) To minimize the model's loss function
- D) To handle missing data

Answer: A) To measure the distance between two probability distributions

Question 78: In Data Science, what is the process of transforming data into a more suitable representation for analysis called?

- A) Data preprocessing
- B) Data augmentation
- C) Data integration
- D) Data wrangling

Answer: A) Data preprocessing

Question 79: Which machine learning algorithm is used for recommending items to users based on their preferences?

- A) Decision Trees
- B) K-Nearest Neighbors (KNN)
- C) Collaborative Filtering
- D) Naive Bayes

Answer: C) Collaborative Filtering

Question 80: What is the process of converting categorical data into numerical representations by assigning integer labels called?

- A) Tokenization
- B) One-hot encoding
- C) Lemmatization
- D) Label encoding

Answer: D) Label encoding

Question 81: In machine learning, what is the process of dividing a dataset into training, validation, and testing sets called?

- A) Data preprocessing
- B) Data augmentation
- C) Data splitting
- D) Data transformation

Answer: C) Data splitting

Question 82: Which Python library is commonly used for data visualization with a focus on high-level interface and attractive aesthetics?

- A) Matplotlib
- B) Seaborn
- C) Pandas
- D) NumPy

Answer: B) Seaborn

Question 83: What is the primary purpose of the activation function in a neural network?

- A) To compute the loss function
- B) To reduce overfitting
- C) To compute the gradient descent
- D) To introduce non-linearity in the model

Answer: D) To introduce non-linearity in the model

Question 84: In Data Science, what is the process of transforming data into a standard format for analysis called?

- A) Data integration
- B) Data preprocessing
- C) Data augmentation
- D) Data scaling

Answer: B) Data preprocessing

Question 85: Which machine learning algorithm is used for text generation and language translation tasks?

- A) Decision Trees



B) Recurrent Neural Networks (RNN)

C) K-Means

D) Logistic Regression

Answer: B) Recurrent Neural Networks (RNN)

Question 86: What is the primary purpose of the F1-score in machine learning evaluation?

A) To measure the accuracy of the model

B) To measure the precision of the model

C) To measure the recall of the model

D) To balance precision and recall in binary classification

Answer: D) To balance precision and recall in binary classification

Question 87: In Data Science, what is the process of transforming text data into numerical vectors with term frequency and inverse document frequency called?

A) Tokenization

B) Lemmatization

C) Term Frequency-Inverse Document Frequency (TF-IDF)

D) Bag-of-Words

Answer: C) Term Frequency-Inverse Document Frequency