

Basic Statistics

1. What is the mean of the following data set: 2, 5, 8, 10, 12?

- A) 5.5
- B) 7.4
- C) 7
- D) 6.5

Answer: C) 7

2. Which of the following is not a measure of central tendency?

- A) Mean
- B) Median
- C) Mode
- D) Range

Answer: D) Range

3. The median is:

- A) The value that appears most often in the data
- B) The middle value when data is arranged in ascending or descending order
- C) The average of the data
- D) The sum of all values divided by the number of values

Answer: B) The middle value when data is arranged in ascending or descending order

4. Which measure of spread is most affected by extreme values?

- A) Range
- B) Interquartile Range
- C) Variance
- D) Standard Deviation

Answer: A) Range

5. The sum of the deviations of a data set from its mean is:

- A) Zero
- B) Equal to the variance
- C) Equal to the standard deviation
- D) Equal to the mode

Answer: A) Zero

6. Which of the following is a characteristic of a normal distribution?

- A) It is skewed to the left
- B) It is symmetric around the mean
- C) It has a uniform distribution
- D) It is always bimodal

Answer: B) It is symmetric around the mean

7. If the standard deviation is zero, the data points are:

- A) Very dispersed
- B) Very similar or identical
- C) Spread out evenly
- D) Randomly distributed

Answer: B) Very similar or identical

8. In a normal distribution, approximately what percentage of the data lies within one standard deviation of the mean?

- A) 50%
- B) 68%
- C) 95%

- D) 99%

Answer: B) 68%

9. What does a z-score represent?

- A) The number of standard deviations a data point is from the mean
- B) The sum of squared deviations
- C) The correlation between two variables
- D) The proportion of values above the mean

Answer: A) The number of standard deviations a data point is from the mean

10. Which of the following is an example of a discrete variable?

- A) Height
- B) Weight
- C) Number of children in a family
- D) Time

Answer: C) Number of children in a family

ANOVA (Analysis of Variance)

11. What is the main purpose of ANOVA?

- A) To compare the mean of two populations
- B) To compare the means of more than two groups
- C) To measure the correlation between two variables
- D) To estimate the population mean

Answer: B) To compare the means of more than two groups

12. Which of the following assumptions is required for ANOVA?

- A) The data is normally distributed
- B) The samples are independent
- C) The variances are equal across groups
- D) All of the above

Answer: D) All of the above

13. In ANOVA, the null hypothesis states that:

- A) All group means are equal
- B) At least one group mean is different
- C) The variance within groups is equal to the variance between groups
- D) None of the above

Answer: A) All group means are equal

14. What does the F-statistic in ANOVA represent?

- A) The ratio of between-group variance to within-group variance
- B) The ratio of the sample mean to the population mean
- C) The difference between the highest and lowest group means
- D) The sum of squared deviations

Answer: A) The ratio of between-group variance to within-group variance

15. If the p-value in an ANOVA test is less than 0.05, you:

- A) Fail to reject the null hypothesis
- B) Reject the null hypothesis
- C) Accept the null hypothesis
- D) Cannot make a decision

Answer: B) Reject the null hypothesis

16. Which of the following is a limitation of ANOVA?

- A) It only compares two groups
- B) It assumes that all groups have the same variance
- C) It cannot be used for continuous data
- D) It is not applicable for normally distributed data

Answer: B) It assumes that all groups have the same variance

17. A significant result in ANOVA means that:

- A) The sample means are equal
- B) There is evidence to suggest that the group means differ
- C) The data has no variance
- D) The null hypothesis is true

Answer: B) There is evidence to suggest that the group means differ

18. Post-hoc tests in ANOVA are used to:

- A) Test assumptions of normality
- B) Compare means of all pairs of groups after a significant ANOVA result
- C) Estimate the variance between groups
- D) Compute the F-statistic

Answer: B) Compare means of all pairs of groups after a significant ANOVA result

19. What is the primary difference between one-way and two-way ANOVA?

- A) One-way ANOVA compares two groups, while two-way compares more than two
- B) One-way ANOVA involves one independent variable, while two-way involves two
- C) Two-way ANOVA is more difficult to compute
- D) One-way ANOVA is used for dependent variables, while two-way is used for independent variables

Answer: B) One-way ANOVA involves one independent variable, while two-way involves two

20. Which of the following is true about the F-distribution in ANOVA?

- A) It is symmetric
- B) It has a mean of zero
- C) It is skewed to the right
- D) It is normally distributed

Answer: C) It is skewed to the right

Hypothesis Testing

21. The null hypothesis (H0) in a hypothesis test represents:

- A) The assumption that there is a significant effect or relationship
- B) The assumption that there is no significant effect or relationship
- C) The observed outcome of the study
- D) A general theory to be tested

Answer: B) The assumption that there is no significant effect or relationship

22. A p-value less than 0.05 indicates:

- A) Strong evidence against the null hypothesis
- B) Weak evidence against the null hypothesis
- C) Evidence in favor of the null hypothesis

- D) That the test is inconclusive

Answer: A) Strong evidence against the null hypothesis

23. A Type I error occurs when:

- A) You fail to reject the null hypothesis when it is actually false
- B) You reject the null hypothesis when it is actually true
- C) The sample size is too small
- D) The p-value is too large

Answer: B) You reject the null hypothesis when it is actually true

24. A Type II error occurs when:

- A) You fail to reject the null hypothesis when it is actually false
- B) You reject the null hypothesis when it is actually true
- C) The sample size is too large
- D) The p-value is too small

Answer: A) You fail to reject the null hypothesis when it is actually false

25. In a two-tailed test, the rejection region is:

- A) Only in the right tail
- B) Only in the left tail
- C) In both tails of the distribution
- D) In the center of the distribution

Answer: C) In both tails of the distribution

26. If the confidence interval for a population mean does not include zero, what does this suggest?

- A) The null hypothesis should be rejected
- B) The population mean is equal to zero
- C) The sample size is too small
- D) The data is not normally distributed

Answer: A) The null hypothesis should be rejected

27. Which of the following is true about the power of a hypothesis test?

- A) It is the probability of making a Type I error
- B) It is the probability of correctly rejecting the null hypothesis when it is false
- C) It is the same as the sample size
- D) It decreases as the sample size increases

Answer: B) It is the probability of correctly rejecting the null hypothesis when it is false

28. The level of significance (alpha) in hypothesis testing is:

- A) The probability of making a Type II error
- B) The threshold for rejecting the null hypothesis
- C) The sample size
- D) The probability of making a Type I error

Answer: D) The probability of making a Type I error

29. Which of the following is NOT a reason to reject the null hypothesis?

- A) The p-value is less than the significance level
- B) The test statistic is in the critical region
- C) The confidence interval includes zero
- D) The data provides strong evidence against the null hypothesis

Answer: C) The confidence interval includes zero

30. Which test would be appropriate for comparing the means of two independent samples?

- A) Paired t-test
- B) One-way ANOVA

- C) Two-sample t-test
 - D) Chi-square test
- Answer:** C) Two-sample t-test

Probability

31. The probability of an event occurring is always between:

- A) -1 and 1
- B) 0 and 1
- C) 0 and 100
- D) -1 and 0

Answer: B) 0 and 1

32. If two events A and B are mutually exclusive, what is $P(A \text{ and } B)$?

- A) $P(A) + P(B)$
- B) $P(A) \times P(B)$
- C) 0
- D) 1

Answer: C) 0

33. The probability of the complement of an event A is:

- A) $P(A)$
- B) $1 - P(A)$
- C) $P(A) + P(A')$
- D) 1

Answer: B) $1 - P(A)$

34. Which of the following is an example of independent events?

- A) Rolling a die and drawing a card
- B) Drawing two cards from the same deck
- C) Flipping a coin twice and getting heads both times
- D) Two people pulling the same card from different decks

Answer: A) Rolling a die and drawing a card

35. The probability of rolling a 4 on a fair six-sided die is:

- A) $1/6$
- B) $1/2$
- C) $1/3$
- D) $1/4$

Answer: A) $1/6$

36. If $P(A) = 0.3$ and $P(B) = 0.5$, what is the probability that either event A or B occurs, assuming A and B are mutually exclusive?

- A) 0.8
- B) 0.5
- C) 0.3
- D) 0.15

Answer: A) 0.8

37. What is the probability of getting heads on two independent flips of a fair coin?

- A) $1/2$
- B) $1/4$
- C) $1/3$

- D) $1/8$

Answer: B) $1/4$

38. The expected value of a random variable is:

- A) The most likely value
- B) The average value weighted by the probabilities of each outcome
- C) The sum of all values in the distribution
- D) Always equal to zero

Answer: B) The average value weighted by the probabilities of each outcome

39. In a normal distribution, approximately 95% of the data lies within:

- A) 1 standard deviation of the mean
- B) 2 standard deviations of the mean
- C) 3 standard deviations of the mean
- D) 4 standard deviations of the mean

Answer: B) 2 standard deviations of the mean

40. Which of the following is NOT a type of probability distribution?

- A) Normal distribution
- B) Uniform distribution
- C) Poisson distribution
- D) Linear distribution

Answer: D) Linear distribution

Additional Questions

41. What does a continuous probability distribution represent?

- A) A distribution for discrete random variables
- B) A distribution where outcomes can take on any value within a range
- C) A distribution for dependent variables
- D) A distribution with only two outcomes

Answer: B) A distribution where outcomes can take on any value within a range

42. Which of the following is an example of a discrete probability distribution?

- A) Normal distribution
- B) Exponential distribution
- C) Binomial distribution
- D) Uniform distribution

Answer: C) Binomial distribution

43. The law of total probability is used to calculate:

- A) The sum of probabilities across multiple events
- B) The likelihood of different events occurring in sequence
- C) The probability of an event given another event
- D) The complement of an event

Answer: A) The sum of probabilities across multiple events

44. In a probability tree diagram, the branches represent:

- A) The possible outcomes of each event
- B) The events that are dependent
- C) The cumulative probability of events
- D) The mean of the distribution

Answer: A) The possible outcomes of each event

45. What does it mean if events A and B are independent?

- A) The occurrence of one event affects the probability of the other
- B) The occurrence of one event has no effect on the probability of the other
- C) The events are mutually exclusive
- D) The events must occur together

Answer: B) The occurrence of one event has no effect on the probability of the other

46. Which distribution is used to model the number of occurrences of an event in a fixed interval of time or space?

- A) Normal distribution
- B) Poisson distribution
- C) Exponential distribution
- D) Binomial distribution

Answer: B) Poisson distribution

47. In a binomial distribution, what does the parameter 'n' represent?

- A) The probability of success
- B) The number of trials
- C) The number of successes
- D) The expected value

Answer: B) The number of trials

48. What is the variance of a uniform distribution with a range from 0 to 10?

- A) 5
- B) 25
- C) 8.33
- D) 50

Answer: C) 8.33

49. The expected value for a discrete random variable is calculated by:

- A) Adding all possible outcomes
- B) Multiplying each outcome by its probability
- C) Subtracting the mean from each outcome
- D) Dividing the variance by the number of outcomes

Answer: B) Multiplying each outcome by its probability

50. What is the probability of drawing a red card from a standard deck of cards?

- A) $1/2$
- B) $1/4$
- C) $1/13$
- D) $1/26$

Answer: A) $1/2$

Python Basics

1. Which of the following is the correct syntax to print "Hello, World!" in Python?
 - A) print("Hello, World!")
 - B) echo("Hello, World!")
 - C) printf("Hello, World!")
 - D) print("Hello World!")**Answer: A) print("Hello, World!")**
2. What will be the output of the following code? `print(3 * 'hello')`
 - A) hellohellohello
 - B) hello3
 - C) 3hello
 - D) hellohellohellohello**Answer: A) hellohellohello**
3. How do you insert a comment in Python?
 - A) // This is a comment
 - B) # This is a comment
 - C) /* This is a comment */
 - D) -- This is a comment**Answer: B) # This is a comment**
4. Which of the following is used to define a function in Python?
 - A) def
 - B) func
 - C) function
 - D) define**Answer: A) def**
5. Which of the following data types is immutable in Python?
 - A) List
 - B) Set
 - C) Dictionary
 - D) Tuple**Answer: D) Tuple**
6. Which of the following is a valid Python variable name?
 - A) 2var
 - B) var_name
 - C) @var
 - D) var-name**Answer: B) var_name**
7. What will be the output of the following Python code? `print(type([1, 2, 3]))`
 - A) <class 'int'>
 - B) <class 'list'>
 - C) <type 'list'>
 - D) <list>**Answer: B) <class 'list'>**
8. How do you create a list in Python?
 - A) list()
 - B) []
 - C) {}

- D) ()

Answer: B) []

9. Which method is used to add an item to a list in Python?

- A) add()
- B) append()
- C) insert()
- D) extend()

Answer: B) append()

10. What does the len() function do in Python?

- A) Returns the last element of a list
- B) Returns the size of the list
- C) Returns the type of an element
- D) Returns the first element of the list

Answer: B) Returns the size of the list

Data Structures

11. Which Python data structure is ordered, changeable, and allows duplicate elements?

- A) Set
- B) List
- C) Dictionary
- D) Tuple

Answer: B) List

12. Which of the following is the correct syntax to create a dictionary in Python?

- A) {key1: value1, key2: value2}
- B) [key1: value1, key2: value2]
- C) (key1: value1, key2: value2)
- D) key1: value1, key2: value2

Answer: A) {key1: value1, key2: value2}

13. Which method is used to remove an item from a dictionary by key?

- A) remove()
- B) pop()
- C) del()
- D) discard()

Answer: B) pop()

14. What is the output of set([1, 1, 2, 3, 3])?

- A) [1, 1, 2, 3]
- B) {1, 2, 3}
- C) (1, 1, 2, 3)
- D) {1, 1, 2, 3, 3}

Answer: B) {1, 2, 3}

15. What will the following code return? list(range(3, 7))

- A) [3, 4, 5, 6, 7]
- B) [3, 4, 5, 6]
- C) [0, 1, 2, 3, 4, 5]
- D) [3, 4, 5, 6, 7, 8]

Answer: B) [3, 4, 5, 6]

16. Which of the following operations can be performed on a set?

- A) Duplicate elements
- B) Indexed elements
- C) Union and intersection
- D) Ordered elements

Answer: C) Union and intersection

17. Which method is used to add an item to a set in Python?

- A) append()
- B) add()
- C) insert()
- D) extend()

Answer: B) add()

18. Which method is used to combine two lists in Python?

- A) add()
- B) concat()
- C) append()
- D) extend()

Answer: D) extend()

19. How do you access the value of a key in a dictionary in Python?

- A) dictionary[key]
- B) dictionary.get(key)
- C) Both A and B
- D) dictionary.key

Answer: C) Both A and B

20. Which function is used to sort a list in Python?

- A) order()
- B) sort()
- C) arrange()
- D) reverse()

Answer: B) sort()

Python Libraries for Data Science

21. Which library is commonly used for data manipulation in Python?

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

Answer: B) Pandas

22. Which of the following is used for numerical computing in Python?

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

Answer: B) NumPy

23. What does the Pandas method `read_csv()` do?

- A) Reads a CSV file into a DataFrame
- B) Writes data to a CSV file

- C) Filters a CSV file
 - D) Converts data into CSV format
 - Answer:** A) Reads a CSV file into a DataFrame
24. Which of the following is used for data visualization in Python?
- A) NumPy
 - B) Pandas
 - C) Matplotlib
 - D) Scikit-learn
- Answer:** C) Matplotlib
25. Which method in Pandas is used to view the first few rows of a DataFrame?
- A) first()
 - B) head()
 - C) tail()
 - D) preview()
- Answer:** B) head()
26. What does the `axis=0` parameter in Pandas refer to?
- A) Operations along columns
 - B) Operations along rows
 - C) Operations along diagonal
 - D) Operations along values
- Answer:** B) Operations along rows
27. Which Python library is best for machine learning?
- A) Pandas
 - B) NumPy
 - C) Scikit-learn
 - D) Matplotlib
- Answer:** C) Scikit-learn
28. What is the primary purpose of the `groupby()` function in Pandas?
- A) Group data by column
 - B) Perform statistical operations on grouped data
 - C) Filter data based on groups
 - D) Merge multiple DataFrames
- Answer:** B) Perform statistical operations on grouped data
29. What is the difference between NumPy arrays and lists in Python?
- A) NumPy arrays are immutable
 - B) Lists are more efficient for large data
 - C) NumPy arrays support element-wise operations
 - D) Lists support multidimensional data
- Answer:** C) NumPy arrays support element-wise operations
30. Which Python library is best for deep learning?
- A) Matplotlib
 - B) TensorFlow
 - C) Pandas
 - D) Scikit-learn
- Answer:** B) TensorFlow
-

31. What is overfitting in a machine learning model?

- A) The model is too simple
- B) The model is too complex and captures noise
- C) The model is underperforming
- D) The model has too few parameters

Answer: B) The model is too complex and captures noise

32. What is the purpose of the `train_test_split()` function in Scikit-learn?

- A) Split the data into training and test sets
- B) Evaluate the model performance
- C) Normalize the data
- D) Perform cross-validation

Answer: A) Split the data into training and test sets

33. Which of the following is used to evaluate classification models?

- A) Mean Squared Error
- B) Confusion Matrix
- C) R-squared
- D) AUC-ROC Curve

Answer: B) Confusion Matrix

34. What is a hyperparameter in machine learning?

- A) A parameter learned from the data
- B) A parameter set before training the model
- C) A model's output
- D) A feature used in the model

Answer: B) A parameter set before training the model

35. What does the K-means algorithm do?

- A) Classifies data into predefined categories
- B) Finds a linear relationship between variables
- C) Groups data into clusters based on similarity
- D) Reduces the dimensionality of data

Answer: C) Groups data into clusters based on similarity

36. Which machine learning algorithm is used for regression tasks?

- A) K-Nearest Neighbors
- B) Decision Trees
- C) Linear Regression
- D) Support Vector Machine

Answer: C) Linear Regression

37. Which Python library is used for statistical analysis?

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

Answer: A) SciPy

38. What is the purpose of regularization in machine learning?

- A) To increase the complexity of the model
- B) To prevent overfitting
- C) To decrease the number of features
- D) To split data into training and test sets

Answer: B) To prevent overfitting

39. Which of the following is an example of a supervised learning algorithm?

- A) K-Means Clustering

- B) PCA (Principal Component Analysis)
- C) Linear Regression
- D) DBSCAN

Answer: C) Linear Regression

40. What is the purpose of the `fit()` method in Scikit-learn?

- A) To train the model on data
- B) To make predictions on data
- C) To evaluate the model performance
- D) To split data into training and test sets

Answer: A) To train the model on data

Advanced Topics

41. What does PCA (Principal Component Analysis) do?

- A) Reduces the dimensionality of data
- B) Splits the data into different groups
- C) Visualizes data in 2D
- D) Identifies outliers in data

Answer: A) Reduces the dimensionality of data

42. What does the `cross_val_score()` function do in Scikit-learn?

- A) Evaluates the performance of a model using cross-validation
- B) Splits the data into training and validation sets
- C) Computes the accuracy of a model
- D) Fits a model to data

Answer: A) Evaluates the performance of a model using cross-validation

43. Which of the following is used to prevent multicollinearity in regression models?

- A) Regularization
- B) Feature Scaling
- C) Cross-validation
- D) Principal Component Analysis

Answer: A) Regularization

44. Which algorithm is used for dimensionality reduction?

- A) K-Means
- B) PCA
- C) Linear Regression
- D) Decision Trees

Answer: B) PCA

45. Which of the following is used to evaluate regression models?

- A) Precision
- B) Accuracy
- C) Mean Squared Error
- D) F1 Score

Answer: C) Mean Squared Error

46. In deep learning, what is the role of the activation function?

- A) It normalizes the data
- B) It transforms the output of neurons to a desired range
- C) It reduces the dimensionality

- D) It splits data into training and test sets

Answer: B) It transforms the output of neurons to a desired range

47. What is a neural network used for?

- A) Classification and regression tasks
- B) Dimensionality reduction
- C) Clustering data
- D) Feature scaling

Answer: A) Classification and regression tasks

48. Which of the following is an ensemble method in machine learning?

- A) Linear Regression
- B) Random Forest
- C) K-Nearest Neighbors
- D) Logistic Regression

Answer: B) Random Forest

49. Which method is used to find the importance of features in a Random Forest model?

- A) fit()
- B) feature_importances_
- C) score()
- D) predict()

Answer: B) feature_importances_

50. Which Python library is typically used for natural language processing (NLP)?

- A) TensorFlow
- B) Pandas
- C) NLTK
- D) SciPy

Answer: C) NLTK

Python Basics

1. **Which of the following is the correct syntax to print "Hello, World!" in Python?**
 - A) print("Hello, World!")
 - B) echo("Hello, World!")
 - C) printf("Hello, World!")
 - D) print("Hello World!")**Answer: A) print("Hello, World!")**
2. **What will be the output of the following code? `print(3 * 'hello')`**
 - A) hellohellohello
 - B) hello3
 - C) 3hello
 - D) hellohellohellohello**Answer: A) hellohellohello**
3. **How do you insert a comment in Python?**
 - A) // This is a comment
 - B) # This is a comment
 - C) /* This is a comment */
 - D) -- This is a comment**Answer: B) # This is a comment**
4. **Which of the following is used to define a function in Python?**
 - A) def
 - B) func
 - C) function
 - D) define**Answer: A) def**
5. **Which of the following data types is immutable in Python?**
 - A) List
 - B) Set
 - C) Dictionary
 - D) Tuple**Answer: D) Tuple**
6. **Which of the following is a valid Python variable name?**
 - A) 2var
 - B) var_name
 - C) @var
 - D) var-name**Answer: B) var_name**
7. **What will be the output of the following Python code? `print(type([1, 2, 3]))`**
 - A) <class 'int'>
 - B) <class 'list'>
 - C) <type 'list'>
 - D) <list>**Answer: B) <class 'list'>**
8. **How do you create a list in Python?**
 - A) list()
 - B) []
 - C) {}
 - D) ()**Answer: B) []**

9. Which method is used to add an item to a list in Python?

- A) add()
- B) append()
- C) insert()
- D) extend()

Answer: B) append()

10. What does the `len()` function do in Python?

- A) Returns the last element of a list
- B) Returns the size of the list
- C) Returns the type of an element
- D) Returns the first element of the list

Answer: B) Returns the size of the list

Data Structures

11. Which Python data structure is ordered, changeable, and allows duplicate elements?

- A) Set
- B) List
- C) Dictionary
- D) Tuple

Answer: B) List

12. Which of the following is the correct syntax to create a dictionary in Python?

- A) `{key1: value1, key2: value2}`
- B) `[key1: value1, key2: value2]`
- C) `(key1: value1, key2: value2)`
- D) `key1: value1, key2: value2`

Answer: A) `{key1: value1, key2: value2}`

13. Which method is used to remove an item from a dictionary by key?

- A) remove()
- B) pop()
- C) del()
- D) discard()

Answer: B) pop()

14. What is the output of `set([1, 1, 2, 3, 3])`?

- A) `[1, 1, 2, 3]`
- B) `{1, 2, 3}`
- C) `(1, 1, 2, 3)`
- D) `{1, 1, 2, 3, 3}`

Answer: B) `{1, 2, 3}`

15. What will the following code return? `list(range(3, 7))`

- A) `[3, 4, 5, 6, 7]`
- B) `[3, 4, 5, 6]`
- C) `[0, 1, 2, 3, 4, 5]`
- D) `[3, 4, 5, 6, 7, 8]`

Answer: B) `[3, 4, 5, 6]`

16. Which of the following operations can be performed on a set?

- A) Duplicate elements
- B) Indexed elements

- C) Union and intersection
- D) Ordered elements

Answer: C) Union and intersection

17. Which method is used to add an item to a set in Python?

- A) append()
- B) add()
- C) insert()
- D) extend()

Answer: B) add()

18. Which method is used to combine two lists in Python?

- A) add()
- B) concat()
- C) append()
- D) extend()

Answer: D) extend()

19. How do you access the value of a key in a dictionary in Python?

- A) dictionary[key]
- B) dictionary.get(key)
- C) Both A and B
- D) dictionary.key

Answer: C) Both A and B

20. Which function is used to sort a list in Python?

- A) order()
- B) sort()
- C) arrange()
- D) reverse()

Answer: B) sort()

Python Libraries for Data Science

21. Which library is commonly used for data manipulation in Python?

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

Answer: B) Pandas

22. Which of the following is used for numerical computing in Python?

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

Answer: B) NumPy

23. What does the Pandas method `read_csv()` do?

- A) Reads a CSV file into a DataFrame
- B) Writes data to a CSV file
- C) Filters a CSV file
- D) Converts data into CSV format

Answer: A) Reads a CSV file into a DataFrame

24. Which of the following is used for data visualization in Python?

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

Answer: C) Matplotlib

25. Which method in Pandas is used to view the first few rows of a DataFrame?

- A) first()
- B) head()
- C) tail()
- D) preview()

Answer: B) head()

26. What does the `axis=0` parameter in Pandas refer to?

- A) Operations along columns
- B) Operations along rows
- C) Operations along diagonal
- D) Operations along values

Answer: B) Operations along rows

27. Which Python library is best for machine learning?

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

Answer: C) Scikit-learn

28. What is the primary purpose of the `groupby()` function in Pandas?

- A) Group data by column
- B) Perform statistical operations on grouped data
- C) Filter data based on groups
- D) Merge multiple DataFrames

Answer: B) Perform statistical operations on grouped data

29. What is the difference between NumPy arrays and lists in Python?

- A) NumPy arrays are immutable
- B) Lists are more efficient for large data
- C) NumPy arrays support element-wise operations
- D) Lists support multidimensional data

Answer: C) NumPy arrays support element-wise operations

30. Which Python library is best for deep learning?

- A) Matplotlib
- B) TensorFlow
- C) Pandas
- D) Scikit-learn

Answer: B) TensorFlow

Data Science Concepts

31. What is overfitting in a machine learning model?

- A) The model is too simple
- B) The model is too complex and captures noise
- C) The model is underperforming

- D) The model has too few parameters

Answer: B) The model is too complex and captures noise

32. What is the purpose of the `train_test_split()` function in Scikit-learn?

- A) Split the data into training and test sets
- B) Evaluate the model performance
- C) Normalize the data
- D) Perform cross-validation

Answer: A) Split the data into training and test sets

33. Which of the following is used to evaluate classification models?

- A) Mean Squared Error
- B) Confusion Matrix
- C) R-squared
- D) AUC-ROC Curve

Answer: B) Confusion Matrix

34. What is a hyperparameter in machine learning?

- A) A parameter learned from the data
- B) A parameter set before training the model
- C) A model's output
- D) A feature used in the model

Answer: B) A parameter set before training the model

35. What does the K-means algorithm do?

- A) Classifies data into predefined categories
- B) Finds a linear relationship between variables
- C) Groups data into clusters based on similarity
- D) Reduces the dimensionality of data

Answer: C) Groups data into clusters based on similarity

36. Which machine learning algorithm is used for regression tasks?

- A) K-Nearest Neighbors
- B) Decision Trees
- C) Linear Regression
- D) Support Vector Machine

Answer: C) Linear Regression

37. Which Python library is used for statistical analysis?

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

Answer: A) SciPy

38. What is the purpose of regularization in machine learning?

- A) To increase the complexity of the model
- B) To prevent overfitting
- C) To decrease the number of features
- D) To split data into training and test sets

Answer: B) To prevent overfitting

39. Which of the following is an example of a supervised learning algorithm?

- A) K-Means Clustering
- B) PCA (Principal Component Analysis)
- C) Linear Regression
- D) DBSCAN

Answer: C) Linear Regression

40. What is the purpose of the `fit()` method in Scikit-learn?

- A) To train the model on data
- B) To make predictions on data
- C) To evaluate the model performance
- D) To split data into training and test sets

Answer: A) To train the model on data

Advanced Topics

41. What does PCA (Principal Component Analysis) do?

- A) Reduces the dimensionality of data
- B) Splits the data into different groups
- C) Visualizes data in 2D
- D) Identifies outliers in data

Answer: A) Reduces the dimensionality of data

42. What does the `cross_val_score()` function do in Scikit-learn?

- A) Evaluates the performance of a model using cross-validation
- B) Splits the data into training and validation sets
- C) Computes the accuracy of a model
- D) Fits a model to data

Answer: A) Evaluates the performance of a model using cross-validation

43. Which of the following is used to prevent multicollinearity in regression models?

- A) Regularization
- B) Feature Scaling
- C) Cross-validation
- D) Principal Component Analysis

Answer: A) Regularization

44. Which algorithm is used for dimensionality reduction?

- A) K-Means
- B) PCA
- C) Linear Regression
- D) Decision Trees

Answer: B) PCA

45. Which of the following is used to evaluate regression models?

- A) Precision
- B) Accuracy
- C) Mean Squared Error
- D) F1 Score

Answer: C) Mean Squared Error

46. In deep learning, what is the role of the activation function?

- A) It normalizes the data
- B) It transforms the output of neurons to a desired range
- C) It reduces the dimensionality
- D) It splits data into training and test sets

Answer: B) It transforms the output of neurons to a desired range

47. What is a neural network used for?

- A) Classification and regression tasks
- B) Dimensionality reduction
- C) Clustering data

- D) Feature scaling

Answer: A) Classification and regression tasks

48. Which of the following is an ensemble method in machine learning?

- A) Linear Regression
- B) Random Forest
- C) K-Nearest Neighbors
- D) Logistic Regression

Answer: B) Random Forest

49. Which method is used to find the importance of features in a Random Forest model?

- A) fit()
- B) feature_importances_
- C) score()
- D) predict()

Answer: B) feature_importances_

50. Which Python library is typically used for natural language processing (NLP)?

- A) TensorFlow
- B) Pandas
- C) NLTK
- D) SciPy

Answer: C) NLTK

I hope this helps! Let me know if you need any more questions or clarification.

You said:

give me 100 mcqs for python datascience machine learning,deep learning,natural language processing,dataframes with answers

ChatGPT said:

Here is a set of 100 multiple-choice questions (MCQs) covering Python for Data Science, Machine Learning, Deep Learning, Natural Language Processing (NLP), and DataFrames with answers:

Python Basics & Data Structures

1. Which of the following is a correct syntax to print "Hello, World!" in Python?

- A) print("Hello World")
- B) echo("Hello World")
- C) printf("Hello World")
- D) print(Hello World)

Answer: A) print("Hello World")

2. Which data type is mutable in Python?

- A) Tuple
- B) List
- C) String

- D) Integer

Answer: B) List

3. **What is the default value of a Python dictionary key that does not exist?**

- A) None
- B) 0
- C) False
- D) KeyError

Answer: A) None

4. **How do you define a function in Python?**

- A) function name()
 - B) def function_name()
 - C) func function_name()
 - D) define function_name()
- Answer:** B) def function_name()

5. **Which of the following data types can hold multiple values in Python?**

- A) Integer
- B) String
- C) List
- D) Boolean

Answer: C) List

6. **How would you create a Pandas DataFrame?**

- A) df = DataFrame()
- B) df = pandas.DataFrame()
- C) df = pd.create_dataframe()
- D) df = pd.data_frame()

Answer: B) df = pandas.DataFrame()

7. **Which function in Python is used to add an item to the end of a list?**

- A) add()
- B) append()
- C) insert()
- D) extend()

Answer: B) append()

8. **What is the output of len([1, 2, 3]) in Python?**

- A) 3
- B) [1, 2, 3]
- C) 6
- D) None

Answer: A) 3

9. **Which of the following is the correct way to create a tuple?**

- A) tuple = (1, 2, 3)
- B) tuple = [1, 2, 3]
- C) tuple = {1, 2, 3}
- D) tuple = <1, 2, 3>

Answer: A) tuple = (1, 2, 3)

10. **Which of the following is an immutable data type in Python?**

- A) List
- B) Dictionary
- C) Set
- D) Tuple

Answer: D) Tuple

DataFrames (Pandas)

11. How do you import the pandas library in Python?

- A) import pandas as pd
- B) import pandas as pn
- C) from pandas import pd
- D) import pandas

Answer: A) import pandas as pd

12. Which method is used to view the first five rows of a DataFrame in Pandas?

- A) head()
- B) first()
- C) view()
- D) top()

Answer: A) head()

13. How would you select a specific column from a DataFrame?

- A) df.column_name
- B) df[column_name]
- C) df->column_name
- D) df.get(column_name)

Answer: B) df[column_name]

14. How do you drop rows from a DataFrame?

- A) df.drop()
- B) df.remove()
- C) df.del()
- D) df.delete()

Answer: A) df.drop()

15. How do you filter rows in a DataFrame based on a condition?

- A) df.filter()
- B) df[df.column_name > 10]
- C) df.query()
- D) df.select()

Answer: B) df[df.column_name > 10]

16. Which of the following is used to read a CSV file into a DataFrame in Pandas?

- A) pd.read_csv()
- B) pd.load_csv()
- C) pd.csv_read()
- D) pd.import_csv()

Answer: A) pd.read_csv()

17. What does `df.describe()` do in Pandas?

- A) Shows summary statistics of the DataFrame
- B) Shows the first few rows of the DataFrame
- C) Describes the data types in the DataFrame
- D) Removes null values from the DataFrame

Answer: A) Shows summary statistics of the DataFrame

18. Which method is used to get the column names of a DataFrame?

- A) df.columns()
- B) df.columns
- C) df.get_columns()

- D) df.list_columns()

Answer: B) df.columns

19. How do you check for missing values in a DataFrame?

- A) df.isnull()
- B) df.null()
- C) df.missing()
- D) df.check_null()

Answer: A) df.isnull()

20. Which method is used to merge two DataFrames in Pandas?

- A) df.concat()
- B) df.append()
- C) df.merge()
- D) df.join()

Answer: C) df.merge()

Machine Learning

21. Which of the following algorithms is used for classification tasks?

- A) K-Means
- B) Linear Regression
- C) Logistic Regression
- D) PCA

Answer: C) Logistic Regression

22. Which Scikit-learn function is used to split the dataset into training and test sets?

- A) train_test_split()
- B) split_data()
- C) partition_data()
- D) data_split()

Answer: A) train_test_split()

23. What does the fit() method do in Scikit-learn?

- A) Fits the model to the training data
- B) Makes predictions
- C) Evaluates the model
- D) Splits the data

Answer: A) Fits the model to the training data

24. Which of the following is a supervised learning algorithm?

- A) K-Means
- B) DBSCAN
- C) Support Vector Machine (SVM)
- D) PCA

Answer: C) Support Vector Machine (SVM)

25. What does overfitting refer to in machine learning?

- A) Model performs well on unseen data
- B) Model learns too much noise from the training data
- C) Model is underfitting the data
- D) Model cannot generalize to the training data

Answer: B) Model learns too much noise from the training data

26. Which of the following is used to evaluate classification models?

- A) R-squared

- B) Accuracy
- C) Mean Squared Error
- D) Confusion Matrix

Answer: D) Confusion Matrix

27. What is the purpose of the `cross_val_score()` function in Scikit-learn?

- A) Splits data into training and test sets
- B) Evaluates the model using cross-validation
- C) Performs hyperparameter tuning
- D) Makes predictions

Answer: B) Evaluates the model using cross-validation

28. Which of the following algorithms is used for regression tasks?

- A) K-Nearest Neighbors
- B) Decision Trees
- C) Linear Regression
- D) Naive Bayes

Answer: C) Linear Regression

29. What is a hyperparameter in machine learning?

- A) A parameter learned from the data
- B) A parameter set before training the model
- C) The output of the model
- D) A feature used in the model

Answer: B) A parameter set before training the model

30. What does the `score()` function return in Scikit-learn?

- A) The training time
- B) The accuracy of the model
- C) The loss function value
- D) The prediction results

Answer: B) The accuracy of the model

Deep Learning

31. Which of the following is a popular deep learning library in Python?

- A) TensorFlow
- B) Scikit-learn
- C) Pandas
- D) Numpy

Answer: A) TensorFlow

32. What is an epoch in deep learning?

- A) One complete forward and backward pass of all training examples
- B) One layer of a neural network
- C) One training iteration
- D) A single feature in the dataset

Answer: A) One complete forward and backward pass of all training examples

33. Which type of neural network is commonly used for image recognition?

- A) Convolutional Neural Networks (CNNs)
- B) Recurrent Neural Networks (RNNs)
- C) Feedforward Neural Networks (FNNs)
- D) Autoencoders

Answer: A) Convolutional Neural Networks (CNNs)

34. What does the activation function do in a neural network?

- A) It normalizes the input data
- B) It introduces non-linearity to the output of neurons
- C) It reduces the dimensionality of the data
- D) It adjusts the weights of the network

Answer: B) It introduces non-linearity to the output of neurons

35. Which activation function is commonly used in hidden layers of deep neural networks?

- A) Sigmoid
- B) Tanh
- C) ReLU
- D) Softmax

Answer: C) ReLU

36. What is the purpose of dropout in neural networks?

- A) To reduce overfitting by randomly setting some of the weights to zero during training
- B) To increase the number of neurons in the network
- C) To reduce the number of layers in the network
- D) To speed up training

Answer: A) To reduce overfitting by randomly setting some of the weights to zero during training

37. Which of the following is true about a recurrent neural network (RNN)?

- A) RNNs are used to model sequential data
- B) RNNs cannot handle sequences
- C) RNNs are only used for classification tasks
- D) RNNs are only used for image processing

Answer: A) RNNs are used to model sequential data

38. Which of the following is a type of deep learning architecture used for natural language processing?

- A) GAN
- B) CNN
- C) LSTM
- D) KNN

Answer: C) LSTM

39. What is backpropagation in neural networks?

- A) The process of updating weights based on the error from the output
- B) The process of selecting the activation function
- C) The process of splitting data into training and testing sets
- D) The process of normalizing the input data

Answer: A) The process of updating weights based on the error from the output

40. Which optimizer is commonly used in deep learning models?

- A) Gradient Descent
- B) Random Search
- C) Hyperparameter Tuning
- D) Cross-validation

Answer: A) Gradient Descent

Natural Language Processing (NLP)

41. Which library is commonly used for NLP in Python?

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

Answer: B) NLTK

42. What does tokenization mean in NLP?

- A) Removing stop words
- B) Splitting text into smaller units like words or sentences
- C) Calculating sentiment scores
- D) Extracting named entities

Answer: B) Splitting text into smaller units like words or sentences

43. Which of the following is a technique to reduce words to their base form in NLP?

- A) Lemmatization
- B) Stemming
- C) Tokenization
- D) Vectorization

Answer: A) Lemmatization

44. Which model is used for machine translation tasks in NLP?

- A) RNN
- B) Transformer
- C) K-means
- D) Linear Regression

Answer: B) Transformer

45. What is the purpose of the TF-IDF method in NLP?

- A) To remove stop words from a text
- B) To find the sentiment of a document
- C) To calculate the importance of a word in a document
- D) To split a text into tokens

Answer: C) To calculate the importance of a word in a document

46. Which Python library is used for deep learning and NLP tasks like text generation?

- A) Keras
- B) TensorFlow
- C) PyTorch
- D) spaCy

Answer: D) spaCy

47. Which of the following is used to create word embeddings in NLP?

- A) Word2Vec
- B) TF-IDF
- C) Bag of Words
- D) Named Entity Recognition (NER)

Answer: A) Word2Vec

48. What is Named Entity Recognition (NER)?

- A) The process of identifying entities like names, locations, dates, etc., in a text
- B) The process of classifying a text into predefined categories
- C) The process of tokenizing a document
- D) The process of creating word embeddings

Answer: A) The process of identifying entities like names, locations, dates, etc., in a text

49. Which of the following is true about the BERT model in NLP?

- A) BERT is based on a transformer architecture
- B) BERT is used for image recognition
- C) BERT is a classical machine learning model
- D) BERT uses a CNN for feature extraction

Answer: A) BERT is based on a transformer architecture

50. Which algorithm is used for document classification in NLP?

- A) K-means
- B) Naive Bayes
- C) Logistic Regression
- D) Decision Trees

Answer: B) Naive Bayes

Advanced Topics in Data Science

51. Which of the following is used for dimensionality reduction?

- A) PCA (Principal Component Analysis)
- B) KNN (K-Nearest Neighbors)
- C) Decision Trees
- D) SVM (Support Vector Machines)

Answer: A) PCA (Principal Component Analysis)

52. What is the purpose of regularization in machine learning?

- A) To increase model complexity
- B) To prevent overfitting by penalizing large coefficients
- C) To speed up model training
- D) To split data into training and testing sets

Answer: B) To prevent overfitting by penalizing large coefficients

53. Which of the following is used to evaluate a regression model?

- A) Mean Squared Error (MSE)
- B) Confusion Matrix
- C) Accuracy
- D) ROC Curve

Answer: A) Mean Squared Error (MSE)

54. What is the purpose of grid search in machine learning?

- A) To split the data into training and test sets
- B) To tune hyperparameters for a model
- C) To visualize model performance
- D) To create a confusion matrix

Answer: B) To tune hyperparameters for a model

55. Which of the following is a boosting algorithm?

- A) Random Forest
- B) SVM
- C) AdaBoost
- D) K-means

Answer: C) AdaBoost

Matplotlib

1. Which of the following is the correct syntax to import the Matplotlib library?

- A) `import matplotlib`
- B) `import matplotlib.pyplot as plt`
- C) `from matplotlib import plot`
- D) `import matplotlib.pyplot as pd`

Answer: B) `import matplotlib.pyplot as plt`

2. How do you create a basic line plot in Matplotlib?

- A) `plt.plot(x, y)`
- B) `plt.bar(x, y)`
- C) `plt.scatter(x, y)`
- D) `plt.line(x, y)`

Answer: A) `plt.plot(x, y)`

3. Which function is used to display the plot in Matplotlib?

- A) `show_plot()`
- B) `plt.show()`
- C) `plot.show()`
- D) `display.plot()`

Answer: B) `plt.show()`

4. What does the function `plt.xlabel('label')` do in Matplotlib?

- A) Sets the x-axis label
- B) Sets the y-axis label
- C) Sets the title of the plot
- D) Sets the color of the plot

Answer: A) Sets the x-axis label

5. What is the purpose of `plt.grid(True)` in Matplotlib?

- A) Adds gridlines to the plot
- B) Sets the background color
- C) Changes the axis ticks
- D) Displays the legend

Answer: A) Adds gridlines to the plot

6. Which of the following functions is used to create a bar plot in Matplotlib?

- A) `plt.bar()`
- B) `plt.hist()`
- C) `plt.scatter()`
- D) `plt.plot()`

Answer: A) `plt.bar()`

7. What is the default color of the plot in Matplotlib?

- A) Blue
- B) Green
- C) Red
- D) Black

Answer: A) Blue

8. What does the function `plt.legend()` do in Matplotlib?

- A) Adds a legend to the plot
- B) Sets the color of the plot

- C) Changes the axis labels
- D) Adds gridlines to the plot

Answer: A) Adds a legend to the plot

9. Which function is used to create a scatter plot in Matplotlib?

- A) `plt.scatter()`
- B) `plt.plot()`
- C) `plt.bar()`
- D) `plt.hist()`

Answer: A) `plt.scatter()`

10. How do you save a figure to a file in Matplotlib?

- A) `plt.save()`
- B) `plt.savefig()`
- C) `plt.storefig()`
- D) `plt.plotfig()`

Answer: B) `plt.savefig()`

Seaborn

11. Which function is used to create a box plot in Seaborn?

- A) `sns.plot()`
- B) `sns.boxplot()`
- C) `sns.histplot()`
- D) `sns.scatterplot()`

Answer: B) `sns.boxplot()`

12. Which function is used to create a heatmap in Seaborn?

- A) `sns.heatmap()`
- B) `sns.scatterplot()`
- C) `sns.histplot()`
- D) `sns.pairplot()`

Answer: A) `sns.heatmap()`

13. What does the `sns.pairplot()` function do in Seaborn?

- A) Creates pairwise scatter plots for numerical data
- B) Creates a histogram for numerical data
- C) Creates a box plot for categorical data
- D) Creates a density plot for numerical data

Answer: A) Creates pairwise scatter plots for numerical data

14. Which function is used to create a violin plot in Seaborn?

- A) `sns.violin()`
- B) `sns.violinplot()`
- C) `sns.boxplot()`
- D) `sns.barplot()`

Answer: B) `sns.violinplot()`

15. What is the default color palette used in Seaborn?

- A) "dark"
- B) "deep"
- C) "light"

- D) "colorblind"

Answer: B) "deep"

16. How do you set the style of a Seaborn plot to white grid?

- A) `sns.set_style('whitegrid')`
- B) `sns.set('whitegrid')`
- C) `sns.set_style('grid')`
- D) `sns.set('grid')`

Answer: A) `sns.set_style('whitegrid')`

17. Which function in Seaborn is used to create a bar plot?

- A) `sns.barplot()`
- B) `sns.histplot()`
- C) `sns.scatterplot()`
- D) `sns.lineplot()`

Answer: A) `sns.barplot()`

18. Which of the following is used to add a title to a Seaborn plot?

- A) `sns.title('Title')`
- B) `plt.title('Title')`
- C) `sns.set_title('Title')`
- D) `sns.add_title('Title')`

Answer: B) `plt.title('Title')`

19. Which function in Seaborn is used to create a pairwise plot of data?

- A) `sns.pairplot()`
- B) `sns.scatterplot()`
- C) `sns.boxplot()`
- D) `sns.lineplot()`

Answer: A) `sns.pairplot()`

20. How can you set the color of the points in a Seaborn scatter plot?

- A) `sns.scatterplot(x='x', y='y', color='red')`
- B) `sns.scatterplot(x='x', y='y', palette='red')`
- C) `sns.scatterplot(x='x', y='y', colors='red')`
- D) `sns.scatterplot(x='x', y='y', hue='red')`

Answer: A) `sns.scatterplot(x='x', y='y', color='red')`

Power BI

21. Which of the following is a primary function of Power BI?

- A) Data processing
- B) Data visualization
- C) Data storage
- D) Data cleaning

Answer: B) Data visualization

22. Which feature of Power BI allows users to create reports and dashboards?

- A) Power Query
- B) Power Pivot
- C) Power BI Desktop
- D) Power View

Answer: C) Power BI Desktop

23. What is the file extension for Power BI Desktop files?

- A) .pbix
- B) .pbit
- C) .xlsx
- D) .pbids

Answer: A) .pbix

24. In Power BI, which of the following is used to connect to data sources?

- A) Power Query
- B) Power Pivot
- C) Power View
- D) Power Apps

Answer: A) Power Query

25. Which of the following charts is NOT available in Power BI by default?

- A) Line chart
- B) Area chart
- C) Pie chart
- D) Radar chart

Answer: D) Radar chart

26. What is the primary use of Power Query in Power BI?

- A) Data visualization
- B) Data cleaning and transformation
- C) Data modeling
- D) Data storage

Answer: B) Data cleaning and transformation

27. Which of the following is used to filter data on a Power BI report page?

- A) Slicer
- B) Filter pane
- C) Bookmarks
- D) Drillthrough

Answer: A) Slicer

28. Which of the following features in Power BI allows users to explore data interactively?

- A) Drilldown
- B) DAX formulas
- C) Power Query
- D) Power Pivot

Answer: A) Drilldown

29. Which of the following can you NOT create in Power BI?

- A) Charts
- B) Dashboards
- C) Reports
- D) Data science models

Answer: D) Data science models

30. Which of the following visualization types is available in Power BI?

- A) Histogram
- B) Gantt Chart
- C) Word Cloud
- D) Bubble chart

Answer: D) Bubble chart

General Data Visualization

31. Which of the following is the default chart type in Power BI for numerical data?

- A) Line Chart
- B) Pie Chart
- C) Column Chart
- D) Scatter Plot

Answer: C) Column Chart

32. Which of the following chart types is most suitable for showing the relationship between two variables?

- A) Pie chart
- B) Histogram
- C) Scatter plot
- D) Line chart

Answer: C) Scatter plot

33. What does a histogram represent in data visualization?

- A) Categorical data distribution
- B) Distribution of a continuous variable
- C) Time series data
- D) Percentage data

Answer: B) Distribution of a continuous variable

34. Which of the following is an appropriate chart to visualize the distribution of categorical data?

- A) Bar Chart
- B) Histogram
- C) Scatter Plot
- D) Line Chart

Answer: A) Bar Chart

35. Which of the following is used to summarize and show data patterns in Power BI?

- A) Reports
- B) Dashboards
- C) Tiles
- D) Visualizations

Answer: D) Visualizations

36. Which chart in Power BI is used to show trends over time?

- A) Area Chart
- B) Line Chart
- C) Pie Chart
- D) Tree Map

Answer: B) Line Chart

37. What kind of chart is typically used to compare parts of a whole?

- A) Bar Chart
- B) Scatter Plot
- C) Pie Chart
- D) Line Chart

Answer: C) Pie Chart

38. Which of the following is used in Power BI to create calculated columns and measures?

- A) Power Query
- B) Power Pivot
- C) DAX (Data Analysis Expressions)
- D) Power Apps

Answer: C) DAX (Data Analysis Expressions)

39. What is the main purpose of creating a Power BI dashboard?

- A) To store data
- B) To share data insights visually
- C) To clean the data
- D) To perform statistical analysis

Answer: B) To share data insights visually

40. Which of the following is used in Power BI to group data by specific categories?

- A) Drillthrough
- B) Grouping
- C) Slicer
- D) Hierarchies

Answer: B) Grouping

Advanced Visualization

41. In Seaborn, which of the following is used to visualize the distribution of a dataset?

- A) `sns.distplot()`
- B) `sns.barplot()`
- C) `sns.pairplot()`
- D) `sns.boxplot()`

Answer: A) `sns.distplot()`

42. Which of the following plots is best suited for showing the relationship between multiple variables in Seaborn?

- A) Heatmap
- B) Pairplot
- C) Boxplot
- D) Violin plot

Answer: B) Pairplot

43. Which of the following chart types is often used in Power BI to visualize geographic data?

- A) Pie chart
- B) Map visualizations
- C) Bar chart
- D) Line chart

Answer: B) Map visualizations

44. In Power BI, which of the following can be used to show hierarchical data?

- A) Hierarchical Slicer
- B) Treemap
- C) Matrix Visual

- D) Card

Answer: B) Treemap

45. Which Power BI feature allows you to automatically refresh the data in reports?

- A) Scheduled refresh
- B) Real-time data connection
- C) Live dashboard updates
- D) Dataflow

Answer: A) Scheduled refresh

46. What is a common use of a waterfall chart in Power BI?

- A) Showing cumulative values over time
- B) Comparing categories of data
- C) Displaying geographical information
- D) Showing continuous changes in data

Answer: A) Showing cumulative values over time

47. Which of the following is used to filter data based on values within Power BI?

- A) Filter Pane
- B) Drillthrough
- C) Report level filters
- D) Both A and C

Answer: D) Both A and C

48. What kind of visualization is ideal for showing hierarchical categories in Seaborn?

- A) Pairplot
- B) Heatmap
- C) Treeplot
- D) Violin plot

Answer: C) Treeplot

49. Which of the following charts can be used in Power BI to visualize parts of a whole?

- A) Tree map
- B) Line chart
- C) Scatter plot
- D) Histogram

Answer: A) Tree map

50. Which visualization in Seaborn is useful for showing the relationship between two variables along with their distribution?

- A) `sns.regplot()`
- B) `sns.scatterplot()`
- C) `sns.lineplot()`
- D) `sns.histplot()`

Answer: A) `sns.regplot()`

Artificial Intelligence (AI)

1. **What is the main goal of Artificial Intelligence (AI)?**
 - A) To make computers faster
 - B) To enable computers to perform tasks that typically require human intelligence
 - C) To increase the speed of data transmission
 - D) To store data more efficiently

Answer: B) To enable computers to perform tasks that typically require human intelligence
2. **Which of the following is a subfield of AI that focuses on creating models that can make decisions based on data?**
 - A) Robotics
 - B) Natural Language Processing (NLP)
 - C) Machine Learning (ML)
 - D) Computer Vision

Answer: C) Machine Learning (ML)
3. **Which of the following best describes supervised learning in machine learning?**
 - A) The model learns from both labeled and unlabeled data
 - B) The model learns from labeled data to make predictions
 - C) The model learns from data without any labels
 - D) The model learns from data with partially correct labels

Answer: B) The model learns from labeled data to make predictions
4. **Which of the following is NOT an example of machine learning?**
 - A) Predicting customer churn
 - B) Automatically identifying spam emails
 - C) Writing software code
 - D) Image recognition

Answer: C) Writing software code
5. **Which of the following algorithms is commonly used for classification tasks?**
 - A) Linear Regression
 - B) K-Nearest Neighbors (KNN)
 - C) K-Means Clustering
 - D) Principal Component Analysis (PCA)

Answer: B) K-Nearest Neighbors (KNN)
6. **Which of the following best defines deep learning?**
 - A) A subset of machine learning that uses neural networks with many layers
 - B) A technique for clustering large datasets
 - C) A supervised learning algorithm for regression problems
 - D) A way to handle missing data

Answer: A) A subset of machine learning that uses neural networks with many layers
7. **Which technique is commonly used to avoid overfitting in machine learning models?**
 - A) Increasing the learning rate
 - B) Regularization
 - C) Decreasing the dataset size
 - D) Using fewer features

Answer: B) Regularization
8. **Which of the following is an example of unsupervised learning?**

- A) Predicting house prices
 - B) Identifying fraudulent transactions
 - C) Clustering customers based on purchasing behavior
 - D) Classifying email messages as spam or not spam
- Answer:** C) Clustering customers based on purchasing behavior

9. **In reinforcement learning, the agent learns by:**

- A) Observing labeled data
- B) Performing actions and receiving rewards or penalties
- C) Clustering data into groups
- D) Using pre-trained models to make predictions

Answer: B) Performing actions and receiving rewards or penalties

10. **Which of the following is a major challenge of AI systems?**

- A) Large amounts of labeled data
- B) Lack of hardware for computation
- C) Computational cost and time
- D) Lack of human intelligence

Answer: C) Computational cost and time

Big Data Analytics

11. **What is the term for large and complex datasets that traditional data processing tools cannot handle?**

- A) Small data
- B) Big data
- C) Clean data
- D) Structured data

Answer: B) Big data

12. **Which of the following is a key characteristic of big data?**

- A) High volume
- B) Low velocity
- C) Minimal veracity
- D) Small variety

Answer: A) High volume

13. **Which of the following is NOT one of the 3Vs of big data?**

- A) Volume
- B) Velocity
- C) Visualization
- D) Variety

Answer: C) Visualization

14. **Which of the following technologies is most commonly used for distributed storage in big data?**

- A) Hadoop Distributed File System (HDFS)
- B) MySQL
- C) Microsoft SQL Server
- D) Oracle Database

Answer: A) Hadoop Distributed File System (HDFS)

15. **Which of the following is the open-source framework used for processing large datasets in big data?**

- A) Apache Spark
- B) Apache Tomcat
- C) Microsoft Power BI
- D) MySQL

Answer: A) Apache Spark

16. What is the purpose of the MapReduce programming model?

- A) To visualize large datasets
- B) To aggregate data across distributed systems
- C) To store big data in a distributed manner
- D) To clean the data before processing

Answer: B) To aggregate data across distributed systems

17. Which of the following is a NoSQL database used for big data storage?

- A) MySQL
- B) MongoDB
- C) PostgreSQL
- D) Microsoft SQL Server

Answer: B) MongoDB

18. Which of the following is a commonly used data format for storing big data?

- A) CSV
- B) XML
- C) Parquet
- D) JSON

Answer: C) Parquet

19. Which of the following is a method used for handling the velocity aspect of big data?

- A) Batch processing
- B) Stream processing
- C) Data warehousing
- D) Data archiving

Answer: B) Stream processing

20. Which of the following is an open-source data processing engine for big data that can process real-time data?

- A) Apache Kafka
- B) Apache Hadoop
- C) Apache Hive
- D) MongoDB

Answer: A) Apache Kafka

AI and Big Data Analytics Applications

21. Which of the following AI techniques is commonly used for sentiment analysis in big data analytics?

- A) Deep Learning
- B) Natural Language Processing (NLP)
- C) K-Means Clustering
- D) Reinforcement Learning

Answer: B) Natural Language Processing (NLP)

22. Which of the following tools is commonly used for data visualization in big data analytics?

- A) Apache HBase
- B) Tableau
- C) Microsoft Excel
- D) Oracle Database

Answer: B) Tableau

23. Which of the following is an example of using AI in big data analytics?

- A) Predictive maintenance in manufacturing
- B) Querying data from a relational database
- C) Data transformation using SQL
- D) Backing up data in a cloud environment

Answer: A) Predictive maintenance in manufacturing

24. Which big data processing tool is primarily used for SQL-like querying on large datasets in Hadoop?

- A) Apache Pig
- B) Apache Hive
- C) Apache Flink
- D) Apache Spark

Answer: B) Apache Hive

25. Which of the following describes "data mining" in the context of big data analytics?

- A) Cleaning the data for processing
- B) Analyzing large datasets to identify patterns
- C) Storing data in a cloud database
- D) Transforming data into a structured format

Answer: B) Analyzing large datasets to identify patterns

26. Which of the following best describes "predictive analytics" in big data?

- A) Analyzing past data to predict future events or behaviors
- B) Visualizing data to explore trends
- C) Sorting data into categories
- D) Cleaning data for accuracy

Answer: A) Analyzing past data to predict future events or behaviors

27. What is the term for analyzing large sets of social media data to understand public opinion or sentiment?

- A) Web scraping
- B) Text mining
- C) Sentiment analysis
- D) Feature engineering

Answer: C) Sentiment analysis

28. Which of the following is NOT an advantage of using big data analytics?

- A) Better decision-making through insights
- B) Improved customer experience
- C) Faster internet speeds
- D) Optimized operational efficiency

Answer: C) Faster internet speeds

29. Which of the following is a key benefit of AI in big data analytics?

- A) Manual data processing
- B) Automated decision-making based on data patterns
- C) Storing large datasets in a warehouse

- D) Reducing data storage costs

Answer: B) Automated decision-making based on data patterns

30. Which type of AI algorithm is commonly used in big data to detect anomalies in large datasets?

- A) Clustering algorithms
- B) Reinforcement learning
- C) Anomaly detection algorithms
- D) Regression algorithms

Answer: C) Anomaly detection algorithms

AI, Big Data and Technologies

31. Which of the following best describes the role of cloud computing in big data analytics?

- A) Providing infrastructure for scalable data storage and processing
- B) Developing AI models
- C) Cleaning and preparing data for analysis
- D) Visualizing the data in reports

Answer: A) Providing infrastructure for scalable data storage and processing

32. What is the primary function of Apache Spark in big data analytics?

- A) Distributed storage of data
- B) Distributed processing of data
- C) Visualization of data
- D) Management of machine learning models

Answer: B) Distributed processing of data