1. Which of the following methods do we use to find the best fit line for data in Linear Regression? A) Least Square Error B) Maximum Likelihood C) Logarithmic Loss D) Both A and B

Answer -: D) Both A and B

2. Which of the following statement is true about outliers in linear regression? A) Linear regression is sensitive to outliers B) linear regression is not sensitive to outliers C) Can't say D) none of these

Answer -: A) Linear regression is sensitive to outliers

3. A line falls from left to right if a slope is _____? A) Positive B) Negative C) Zero D) Undefined

Answer -: B) Negative

4. Which of the following will have symmetric relation between dependent variable and independent variable? A) Regression B) Correlation C) Both of them D) None of these

Answer -: B) Correlation

5. Which of the following is the reason for over fitting condition? A) High bias and high variance B) Low bias and low variance C) Low bias and high variance D) none of these

Answer -: C) Low bias and high variance

6. If output involves label then that model is called as: A) Descriptive model B) Predictive modal C) Reinforcement learning D) All of the above

Answer -: A) Descriptive model

7. Lasso and Ridge regression techniques belong to _____? A) Cross validation B) Removing outliers C) SMOTE D) Regularization

Answer -: D) Regularization

8. To overcome with imbalance dataset which technique can be used? A) Cross validation B) Regularization C) Kernel D) SMOTE

Answer -: D) SMOTE

9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph? A) TPR and FPR B) Sensitivity and precision C) Sensitivity and Specificity D) Recall and precision

Answer -: A) TPR and FPR

10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less. A) True B) False

Answer -: B) False

11. Pick the feature extraction from below: A) Construction bag of words from a email B) Apply PCA to project high dimensional data C) Removing stop words D) Forward selection

Answer -: B) Apply PCA to project high dimensional data

12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression? A) We don't have to choose the learning rate. B) It becomes slow when number of features is very large. C) We need to iterate. D) It does not make use of dependent variable.

Answer -: A) We don't have to choose the learning rate & C) We need to iterate

13. Explain the term regularization?

Answer -: Regularization in machine learning is like adding a speed bump to an automobile (model) to prevent it from fitting the training data too closely. It prevents the model from getting overly complex, which can cause it to perform well on known routes (training data) but poorly on unknown roads (unseen data). Regularization, by introducing this "speed bump," contributes to the development of a smoother, more generic model that performs better in new conditions.

14. Which particular algorithms are used for regularization?

Answer – (1) Ridge Regression (L2 Norm) (2) Lasso (L1 Norm) (3) Dropout

15. Explain the term error present in linear regression equation?

Answer -: The "error" in a linear regression equation is the difference between what the model predicts and the actual values in the data. It's like the leftover or missing part that the model couldn't perfectly capture. The goal is to make these errors as small as possible by finding the best-fitting line.

16. Which of the following operators is used to calculate remainder in a division? A) # B) & C) % D) $

Answer -: C) %

17. In python 2//3 is equal to? A) 0.666 B) 0 C) 1

Answer -: A) 0.666

18. In python, 6<<2 is equal to? A)36 B)10 C)24 D)45

Answer -: C) 24

19. In python, 6&2 will give which of the following as output? A) 2 B) True C) False D) 0

Answer -: A) 2

20. In python, 6|2 will give which of the following as output? A) 2 B) 4 C) 0 D) 6

Answer -: D) 6

21. What does the finally keyword denotes in python? A) It is used to mark the end of the code B) It encloses the lines of code which will be executed if any error occurs while executing the lines of code in the try block. C) the finally block will be executed no matter if the try block raises an error or not. D) None of the above

Answer -: C) the finally block will be executed no matter if the try block raises an error or not.

22. What does raise keyword is used for in python? A) It is used to raise an exception. B) It is used to define lambda function C) it's not a keyword in python. D) None of the above

Answer -: A) It is used to raise an exception

23. Which of the following is a common use case of yield keyword in python? A) in defining an iterator B) while defining a lambda function C) in defining a generator D) in for loop.

Answer -: C) in defining a generator

24. Which of the following are the valid variable names? A) _abc B) 1abc C) abc2 D) None of the above

Answer -: A) _abc

25. Which of the following are the keywords in python? A) yield B) raise C) look-in D)all of the above

Answer -: A) rasie & B) raise

26. Bernoulli random variables take (only) the values 1 and 0. a) True b) False

Answer -: a) True

27. Which of the following theorem states that the distribution of averages of iid variables, properly normalized, becomes that of a standard normal as the sample size increases? a) Central Limit Theorem b) Central Mean Theorem c) Centroid Limit Theorem d) All of the mentioned

Answer -: b) Central Mean Theorem

28. Which of the following is incorrect with respect to use of Poisson distribution? a) Modeling event/time data b) Modeling bounded count data c) Modeling contingency tables d) All of the mentioned

    Answer -: b) Modeling bounded count data

29. Point out the correct statement. a) The exponent of a normally distributed random variables follows what is called the log- normal distribution b) Sums of normally distributed random variables are again normally distributed even if the variables are dependent c) The square of a standard normal random variable follows what is called chi-squared distribution d) All of the mentioned

Answer -: c) The square of a standard normal random variable follows what is called chi-squared distribution

30. _____ random variables are used to model rates. a) Empirical b) Binomial c) Poisson d) All of the mentioned

Answer -: c) Poisson

31. Usually replacing the standard error by its estimated value does change the CLT. a) True b) False

Answer -: b) false

32. Which of the following testing is concerned with making decisions using data? a) Probability b) Hypothesis c) Causal d) None of the mentioned

Answer -: c) Casual

33. Normalized data are centered at_____and have units equal to standard deviations of the original data. a) 0 b) 5 c) 1 d) 10

Answer -: a) 0

34. Which of the following statement is incorrect with respect to outliers? a) Outliers can have varying degrees of influence b) Outliers can be the result of spurious or real processes c) Outliers cannot conform to the regression relationship d) None of the mentioned

Answer -: c) Outliers cannot conform to the regression relationship

35. What do you understand by the term Normal Distribution?

Answer -: The normal distribution, also known as the Gaussian distribution, is a symmetric probability distribution about the mean, indicating that data near the mean occur more frequently than data distant from the mean.

36. How do you handle missing data? What imputation techniques do you recommend?
37. What is A/B testing?

Answer -: A /B testing, also known as split testing, is a method for comparing two versions of something to see which one performs better. It's commonly used in website and app optimization, but can be applied to any situation where you want to determine the best way to achieve a desired outcome.

38. Is mean imputation of missing data acceptable practice?

Answer -: No, Mean Imputation of missing data is not an acceptable practice since it can ruin your data in many ways like shrinking the data, introduces bias and can at times hide relationships as well.

39. What is linear regression in statistics?

Answer -: Linear regression is a way to figure out the straight line that best fits a set of data points. It helps us understand and predict the relationship between two or more variables.

40. What are the various branches of statistics?

Answer -: Statistics is a broad field with various branches that cater to different aspects of data analysis, interpretation, and decision-making. Branches of the Stats are as follows –

- Descriptive Stats – Involves methods for summarizing and describing the main features of a dataset. Measures like mean, median, mode and Standard Deviation fall under descriptive statistics.
- Inferential Stats – Concerned with making inferences or perdictions about a population based on a sample of data. Regression analysis is an example of Inferential Stats.

41. Write a python program to find the factorial of a number

Answer -:

```
num = int(input(“Enter the number”))
fact = 1
if num < 0:
   print(“Factorial does not exists”)
elif num==0:
   print(“The factorial of 0 is 1”)
else:
   for i in range(1,num+1)
      fact = fact*i
   print(fact)
```

42.