**MACHINE LEARNING**

1. Which of the following methods do we use to find the best fit line for data in Linear Regression?

A) Least Square Error

2. Which of the following statement is true about outliers in linear regression?

A) Linear regression is sensitive to outliers

3. A line falls from left to right if a slope is \_\_\_\_\_\_?

B) Negative

4. Which of the following will have symmetric relation between dependent variable and independent

variable?

C) Both of them

5. Which of the following is the reason for over fitting condition?

C) Low bias and high variance

6. If output involves label then that model is called as:

B) Predictive modal

7. Lasso and Ridge regression techniques belong to \_\_\_\_\_\_\_\_\_?

D) Regularization

8. To overcome with imbalance dataset which technique can be used?

A) Cross validation

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

10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the

curve should be less.

B) False

11. Pick the feature extraction from below:

C) Removing stop words

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.

13. Explain the term regularization?

Regularization is used to check whether our data is over fitting our model by removing the high-valued regression co-efficient. Three types of regularization are LASSO, RIDGE, ELASTIC NET.

14. Which particular algorithms are used for regularization?

1-Normalize the values by substracting the mean and dividing by L2 norm.

2-Check the learning rate (alpha)

3-Fit it with the training data and arrive with the score

4-Compare the regularization score with Regression score

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

The difference between the expected value at a particular time and the actual value is called as term error in linear regression equation.

**STATISTICS**

1. Bernoulli random variables take (only) the values 1 and 0.

a) True

2. 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

3. Which of the following is incorrect with respect to use of Poisson distribution?

b) Modeling bounded count data

4. Point out the correct statement.

d) All of the mentioned

5. \_\_\_\_\_\_ random variables are used to model rates.

c) Poisson

6. Usually replacing the standard error by its estimated value does change the CLT.

b) False

7. Which of the following testing is concerned with making decisions using data?

b) Hypothesis

8. Normalized data are centered at\_\_\_\_\_\_and have units equal to standard deviations of the original data.

a) 0

9. Which of the following statement is incorrect with respect to outliers?

c) Outliers cannot conform to the regression relationship

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

Probability distribution that is symmetric about the mean is called normal distribution. Normal distribution curve is bell shaped.

11. How do you handle missing data? What imputation techniques do you recommend?

1. We can delete to eliminate missing data. For only certain datasets, we can use delete and for certain we cannot do the same.

2. To eliminate missing data, regression analysis can be used.

Data imputation techniques like the below can be used:

Mean, median, mode substitution (Based on scenario)

Regression imputation

Last observation carried forward

Maximum likelihood

12. What is A/B testing?

A/B testing is also known as split-run testing. A/B tests consist of a [randomized experiment](https://en.wikipedia.org/wiki/Randomized_experiment) with two variants, A and B. It includes application of [statistical hypothesis testing](https://en.wikipedia.org/wiki/Statistical_hypothesis_testing) as used in the field of [statistics](https://en.wikipedia.org/wiki/Statistics). Two versions of a single [variable](https://en.wikipedia.org/wiki/Variable_(mathematics)) can be determined using this way.

13. Is mean imputation of missing data acceptable practice?

Yes, in some cases like student marks (where the class of students is highly competitive), among students or the mark of a particular student missing for one test where he attended 5 tests, can be taken into account.

14. What is linear regression in statistics?

Linear regression explains the relationship between one or more independent variable with one outcome. It also shows the impact of multiple independent variables on the dependent variable. We can measure the outcome though the independent variables are on a different scale.

15. What are the various branches of statistics?

The two main branches are descriptive and inferential statistics.

**PYTHON**

1. Which of the following operators is used to calculate remainder in a division?

C) %

2. In python 2//3 is equal to?

B) 0

3. In python, 6<<2 is equal to?

C) 24

4. In python, 6&2 will give which of the following as output?

A) 2

5. In python, 6|2 will give which of the following as output?

D) 6

6. What does the finally keyword denotes in python?

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

7. What does raise keyword is used for in python?

A) It is used to raise an exception.

8. Which of the following is a common use case of yield keyword in python?

A) in defining an iterator

9. Which of the following are the valid variable names?

A) \_abc

C) abc2

10. Which of the following are the keywords in python?

A) yield B) raise

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

12. Write a python program to find whether a number is prime or composite.

13. Write a python program to check whether a given string is palindrome or not.

14. Write a Python program to get the third side of right-angled triangle from two given sides.

15. Write a python program to print the frequency of each of the characters present in a given string.