**Machine Learning**

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

Ans:Least Square Error

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

Ans:Linear regression is sensitive to outliers

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

Ans:Negative

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

Ans:Both of them

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

Ans:Low bias and high variences

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

Ans:Predictive model

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

Ans: Regularization

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

Ans:SMOTE

1. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses \_\_\_\_\_ to make graph?

Ans:TPR and FPR

1. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.

Ans:False

1. Pick the feature extraction from below:

Ans:Apply PCA to project high dimensional data

1. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?

Ans:It becomes slow when number of feature is very large

1. Explain the term regularization?

Ans:1] In mathematics, statistics, finance, computer science, particularly in machine learning and inverse problems, regularization is a process that changes the result answer to be simpler. It is often used to obtain results for ill-posted problems or to prevent. Although regularization procedures can be divided in many ways, the following delineation is particularly helpful.

2]Example of, included; k-means: Restricting the segments for avoiding redundant groups. Neural networks: confining the complexity of a model. Random forest: Reducing the depth of tree and branches.

3]The regularization parameter is control on your fitting parameters.

14] Which particular algorithms are used for regularization?

Ans:1]Ridge regression:- Its purpose is to overcome problems such as data overfitting and multicollinearity in data When there is considerable collinearity the existence of near-linear connections among the independent variables among the feature variables, a typical linear or polynomial regression model will fail. Ridge regression adjust the variables by a modest squared bias factor.

2]LASSO:- It is simply penalize large coefficients, in contrast to ridge regression. When the hyperparameter is big enough, Lasso has the effect of driving certain coefficient estimations to be absolutely zero. If we have a large number of features, LASSO works effectively for feature selection.

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

Ans:- 1]It is the value of y obtained using the regression line. ^y is not equal to y from the data. The term y0-^y0=e0 is called the error.

2]it is often said that the the error term in a regression equation represents the effect of the variables. That were omitted from the equation.

3]An error term represents the margin of error with a statistical model; it refers to the sum of the deviaations within th regression line, which provides an explanation for the difference between the theoretical value of the model and actual observed results.

Python

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

Ans:%

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

Ans:0

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

Ans:24

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

Ans:False

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

Ans:0

6. What does the finally keyword denotes in python?

Ans: It encloses the lines of code which will be executed if any error occurs while executing the lines of code in

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

Ans: It is used to raise an exception.

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

Ans: in for loop.

9. Which of the following are the valid variable names? keywords in python?

Ans:None of these

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

Ans:yield

STATISTICS

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

Ans: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?

Ans:Central limit theorem

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

Ans:Modeling bounded count data

4] Point out the correct statement.

Ans:All of the mentioned

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

Ans:poisson

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

Ans:false

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

Ans:Hypothsis

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

Ans:0

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

Ans:none of the mentioned

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

Ans:1] The normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean.

2]In a normal distribution the mean is zero and the standared deviation is 1. It has zero skew and a kurotosis of 3.

3]In normal distribution is the proper term for probability bell curve

4]Normal distribution are summetrical but not all symmetrical distribution are normal

5]In finance, most pricing distribution are not however, perfectly normal.

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

Ans:1]Deleting rows with missing values

2]impute missing values for continuous variable.

3]Impute missing values for categorical variable

4]other imputation methods.

5]Using algorithms that supports missing values.

4]Prediction of missing values.

5]Imputation is a technique used for replacing the missing data with some substitute value to retain most of the data information of the dataset.

12] What is A/B testing?

Ans:1] A/B testing is known as split testing or bucket testing is a methodology for comparing two versions of webpage or app against each other to determine which one performs better.

2] A/B testing is essentially an experiment where two or more variants of a page are shown to users at random and statistical analysis is used to determine which variation performance better for a given conversion goal.

13] Is mean imputation of missing data acceptable practice?

Ans:1]Mean imputation does not preserve the relationships among variables:-True, imputing the mean preserves the mean of the observed data. So if the data are missing completely at random the estimate of the mean remains unbiased.

2]Mean imputation leads to an underestimate of standared error:-A second reason is applies to any type of single imputation. Any statistic that uses the imputed data will have a standard error of the mean will be too small.

14] What is linear regression in statistics?

Ans:1]Linear regression analysis is used to predict the value of a variable based on the value of another variable.

2]The variable you want to predict is called the dependent variable. The variable you are using to predict the other variables value is called the independent variables

3]linear regression analysis is used to predict the value of a variable based on the value of another variable.

15] What are the various branches of statistics?

Ans:1]The Descriptive statistics and inferential statistics are the two main branches of statistics.

2]both are the two statistics branches are used in scientific data analysis and are equally important for the student of statistics.

3]They represent a measurable quantity.