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

Ans. D) Both A and B

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

Ans. A) Linear regression is sensitive to outliers

A line falls from left to right if a slope is _____?

Ans. B) Negative

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

Ans. C) Both of them

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

Ans. C) Low bias and high variance

If output involves label then that model is called as:

Ans. B) Predictive modal

Lasso and Ridge regression techniques belong to _____?

Ans. D) Regularization

To overcome with imbalance dataset which technique can be used?

Ans. D) SMOTE

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

Ans A) TPR and FPR

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

Ans. B) False

Pick the feature extraction from below:

ANs. A) Construction bag of words from a email

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

Ans. B) It becomes slow when number of features is very large.

Explain the term regularization?

Ans. Regularization is a technique that can be applied to various machine learning algorithms to prevent overfitting and improve generalization

Which particular algorithms are used for regularization?

Ans. Linear Regression, Logistic Regression, Support Vector Machines (SVM), Decision Trees, Neural Networks, Principal Component Analysis (PCA), K-Means Clustering and Elastic Net.

Explain the term error present in linear regression equation?

Ans.

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

Ans. C) %

In python 2//3 is equal to?

Ans. B) 0

In python, 6<<2 is equal to?

Ans. C) 24

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

Ans. D) 0

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

Ans. D) 6

What does the finally keyword denotes in python?

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

What does raise keyword is used for in python?

Ans. A) It is used to raise an exception. B) It is used to define lambda function

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

Ans. C) in defining a generator

Which of the following are the valid variable names?

Ans. A) _abc and C) abc2

Which of the following are the keywords in python?

Ans. A) yield and B) raise

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

Ans. a) True

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

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

b) Modeling bounded count data

. 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

Point out the correct statement.

Ans. c) The square of a standard normal random variable follows what is called chi-squared

distribution

_____ random variables are used to model rates.

Ans c) Poisson

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

Ans. b) False

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

Ans. b) Hypothesis

Normalized data are centered at_____and have units equal to standard deviations of the

original data.

Ans

a) 0

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

Ans. c) Outliers cannot conform to the regression relationship

What do you understand by the term Normal Distribution?

Ans. A normal distribution, also known as a Gaussian distribution or bell curve, is a symmetric probability distribution that is characterized by a bell-shaped curve.

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

Ans. Handling missing data is a critical step in the data preprocessing phase, and the choice of imputation technique depends on the nature of the data and the reasons for missingness.

What is A/B testing?

Ans. A/B testing, also known as split testing, is a statistical method used in marketing, product development, and other fields to compare two versions of a product or strategy and determine which one performs better. The primary goal is to assess the impact of changes and make data-driven decisions

Is mean imputation of missing data acceptable practice?

Ans. Mean imputation, where missing values are replaced with the mean of the observed values in a variable, is a simple and commonly used method for handling missing data.

What is linear regression in statistics?

Ans. Linear regression is a statistical method used for modeling the relationship between a dependent variable and one or more independent variables. The goal of linear regression is to find the best-fitting linear relationship that describes the average change in the dependent variable as a function of changes in the independent variable(s).

What are the various branches of statistics?

Ans. Descriptive Statistics, Inferential Statistics, Biostatistics, Econometrics, Psychometrics, Actuarial Science, Spatial Statistics, Environmental Statistics, Quality Control, Nonparametric Statistics, Statistical Genetics, Social Statistics.