

ASSIGNMENT - 39

MACHINE LEARNING-

Ans 1- option A

Ans 2- option A

Ans 3- option B

Ans 4- option D

Ans 5- option C

Ans 6- option B

Ans 7- option C

Ans 8- option B

Ans 9- option A

Ans 10- option A

Ans 11- option B

Ans 12- option A

Ans 13- Regularization is one of the most important concepts of machine learning. It is a technique to prevent the model from overfitting by adding extra information to it.

The commonly **used regularization techniques are :**

1. Lasso Regularization – L1 Regularization
2. Ridge Regularization – L2 Regularization
3. Elastic Net Regularization – L1 and L2 Regularization.

- Ans 14- **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 adjusts the variables by a modest squared bias factor. The feature variable coefficients are pulled away from this rigidity by such a squared bias factor, providing a little bit of bias into the model but considerably lowering variation.

Ans 15- An **error term** in statistics is a value which represents how observed data differs from actual **population data**. It can also be a variable which represents how a given statistical model differs from reality. The error term is often written ε .

STATISTICS-

Ans 1- option A

Ans 2- option A

Ans 3- option B

Ans 4- option D

Ans 5- option C

Ans 6- option B

Ans 7- option B

Ans 8- option A

Ans 9- option D

Ans 10- 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. In graphical form, the normal distribution appears as a "bell curve".

Ans 11- The handling of missing data is very important during the preprocessing of the dataset as many machine learning algorithms do not support missing values.

This article covers 7 ways to handle missing values in the dataset:

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 support missing values
6. Prediction of missing values
7. Imputation using Deep Learning Library — Datawig

Ans 12- A/B testing, also known as split testing, refers to a randomized experimentation process wherein two or more versions of a variable (web page, page element, etc.) are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drives business metrics.

Ans 13- 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.

Ans 14- Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

Ans 15- Statistics have majorly categorised into two types:

1. Descriptive statistics
2. Inferential statistics

Descriptive Statistics

In this type of statistics, the data is summarised through the given observations. The summarisation is one from a sample of population using parameters such as the mean or [standard deviation](#).

Inferential Statistics

This type of statistics is used to interpret the meaning of Descriptive statistics. That means once the data has been collected, analysed and summarised then we use these stats to describe the meaning of the collected data. Or we can say, it is used to draw conclusions from the data that depends on random variations such as observational errors, sampling variation, etc.