MACHINE LEARNING

Ans 1. (A) Least Square Error

Ans 2. (A)Linear regression is sensitive to outliers

Ans 3. (B)Negative

Ans 4. (A)Regression

Ans 5. (C)Low bias and high variance

Ans 6. (B)Predictive method

Ans 7. (D)Regularization

Ans 8. (D)SMOTE

Ans 9. (C)Sensitivity and Specificity

Ans 10. (B)False

Ans 11. (B)Apply PCA to project high dimensional data

Ans 12. (c) and (D)

Ans 13. Regularization is the process of sort out or minimize the overfitting problem. It minimize the coefficient value towards zero in the model.

Ans 14. LASSO and Ridge are used for regularization.

Ans 15. When the predicted value of y is different from the actual value of y in linear regression called error present in linear regression. It happens because independent value is not always giving outcome as actually which observed earlier.

PYTHON – WORKSHEET 1

Ans 1. (C)%

Ans 2. (B)0

Ans 3. (A)36

Ans 4. (A)2

Ans 5. (D)6

Ans 6. (C)

Ans 7. (A)

Ans 8. (C)

Ans 9. (A,B,C)

Ans 10. (A,B)

STATISTICS WORKSHEET-1

Ans 1. (a)True

Ans 2. (a)Central Limit Theorem

Ans 3. (b)Modeling contingency tables

Ans 4. (d)All of the mentioned

Ans 5. (c)Poisson

Ans 6. (b)False

Ans 7. (b)Hypothesis

Ans 8. (a) 0

Ans 9. (c)

Ans 10. The normal distribution is the most commonly seen continuous distribution in nature. It is also called the Gaussian distribution. Just as the binomial distribution, every event is independent from one another.In the normal distribution, the mean, median and mode all line up such that the center of the distribution is the mean.Because of this, exactly half of the results fall to either side of the mean.The mean is zero and the standard deviation is 1. Skewness is not present here or it is equals to zero. In graph form, normal distribution will appear as a bell curve.

Ans 11. Missing data creates a lot of problem for data analysis because it distorts findings. Here are some best techniques which handle missing data:-

1. Use deletion methods to eliminate missing data

2. Use regression analysis to systematically eliminate data

3. Use scientists can use data imputation techniques

Imputation technique - Imputation is a technique used for replacing the missing data with some substitute value to retain most of the data

of the datasets. It is used because removing the datasets every time is not good or desirable. It may leads to incorrect analysis.

Ans 12. A/B testing is a basically to compare the two versions of a variable to find out which performs better in a controlled environment.

Ans 13. The process of replacing null values in a data collection with the data's mean is know as mean Imputation. Mean imputation is considered terrible practice science it ignores the feature correlation.

Ans 14. Linear regression is a model which shows relationship between two or more variables. Linear regression predicts the value of a variable based on the value of another variable. The variables which is used for predicting the value's of other variable called independent variable and resulting value of variable is called the dependent variable. Example - Study hours is independent variable and marks obtain in exam is variable which depend upon study hours.

Ans 15. The various branches of statistics are following:-

1. Descriptive Statistics - It helps in summarizing and organizing any data set characteristics. It also helps in the representation of data in both classification and diagrammatic way.

2. Inferential Statistics - It helps in finding the conclusion regarding the population after analysis on the sample drawn from it.