# MACHINE LEARNING

1. A) Least Square Error

2. A) Linear regression is sensitive to outliers

3. B) Negative

4. C) Both of them

5. A) High bias and high variance

6. C) Low bias and high variance

7. D) Regularization

8. D) SMOTE

9. A) TPR and FPR

10. B) False

11. B) Apply PCA to project high dimensional data

12. A) We don't have to choose the learning rate.
    B) It becomes slow when number of features is very large.

13. when we use regression model to train some data, there is a good chance that the model will overfitting the given data sets. Regularization is helps sort this overfitting problems. Simply reducing the number of degrees of a polynormal function by reducing their corresponding weights.

Below the different type of Regularization:

1. LASSO Regression(L1 Form):
   LASSO regression Penalizes the model based on sum of magnitude of the co efficient.
   Learning rate should be slow steps, that will be helped make zero error. It will come close to zero error. Its find relationship between feature & target values.
2. Bridge Regression (L2 Form):
   LASSO regression Penalizes the model based on sum of squares of magnitude of the co efficient. It is as similar as LASSO regression.

14. Lasso (L1 Form) & Ridge (L2 Form) are the best algorithm for regularization. LassoCV & RidgeCV.

15. the error term is the difference between the expected price at a particular time and the price that was actually observed.

## STATISTICS WORKSHEET-1

1. a) True

2. a) Central Limit Theorem

3.b) Modeling bounded count data

4.d) All of the mentioned

5. c) Poisson

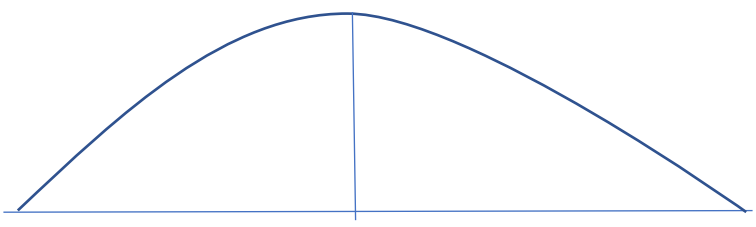
6. b) False

7. b) Hypothesis

8. a) 0

9. c) Outliers cannot conform to the regression relationship

10. Normal 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. The probability distribution that plots all of its values in a symmetrical fashion, and most of the results are situated around the probability's mean.



11. Missing data can be dealt with in a variety of ways. I think that, following step should be check:

1. we have to check how big is the massing data in data set. If it is big then we have to use imputer technique handle this missing data. Missing data is have in continuous data column then we use mean of that column and fill the missing data with mean value. If it is categorical data then we use mode of the column and fill missing value with mode value.

If missing data is small then we can ignore that data. But ignore or deletion is not good way to handle the missing data.

Which missing data is not important. Means, which missing data column is not get more correlation with the target value that column can ignore for analysis.

We will be recommended the following important technique to handle the missing value.

I) Knn imputation

II) Iterative imputation

12. A/B testing is the process of comparing two variations of a page element, usually by testing users' response to variant A vs. variant B and concluding which of the two variants is more effective.

13. Mean imputation is typically considered terrible practice since it ignores feature correlation. Consider the following scenario: we have a table with age and fitness scores, and an eight-year-old has a missing fitness score. If we average the fitness scores of people between the ages of 15 and 80, the eighty-year-old will appear to have a significantly greater fitness level than he actually does.

Second, mean imputation decreases the variance of our data while increasing bias. As a result of the reduced variance, the model is less accurate and the confidence interval is narrower.

14. Linear regression is a basic and commonly used type of predictive analysis. The overall idea of regression is to examine two things: (1) does a set of

predictor variables do a good job in predicting an outcome (dependent) variable? (2) Which variables in particular are significant predictors of the outcome variable, and in what way do they–indicated by the magnitude and sign of the beta estimates–impact the outcome variable? These regression estimates are used to explain the relationship between one dependent variable and one or more independent variables.

15. There are three real branches of statistics.

I) Data Collection

II) Descriptive Statistics: Describe area is call descriptive statistics.

III) Inferential Statistics: Whole data set area is call inferential Statistics. In this, Population & Sample come int picture.

## PYTHON – WORKSHEET 1

1. C) %
2. B) 0
3. C) 24
4. A) 2

5. D) 6
6. C) the finally block will be executed no matter if the try block raises an error or not.
7. A) It is used to raise an exception.
8. A) in defining an iterator
9. A) _abc & C) abc2
10. A) yield & B) raise