**Machine Learning Assignment**

1. Which of the following methods do we use to find the best fit line for data in Linear Regression?
2. Least Square Error
3. Which of the following statement is true about outliers in linear regression?
4. Linear regression is sensitive to outliers

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

B) Negative

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

D) None of these

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

C) Low bias and high variance

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

B) Predictive modal

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

D) Regularization

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

D) SMOTE

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

A) TPR and FPR

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

A) True

11. Pick the feature extraction from below:

C) Removing stop words

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

C) We need to iterate. & D) It does not make use of dependent variable.

14. Which particular algorithms are used for regularization?

**Ans.**

mostly used Ridge Regression :-

Ridge regression is a model tuning method that is used to analyses any data that suffers from multicollinearity. This method performs L2 regularization. When the issue of multicollinearity occurs, least-squares are unbiased, and variances are large, this results in predicted values being far away from the actual values.

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

Ans

The error term is also known as the residual, disturbance, or remainder term, and is variously represented in models by the letters e, ε, or u.

An error term essentially means that the model is not completely accurate and results in differing results during real-world applications. For example, assume there is a multiple linear regression function that takes the following form:

The he two data points with the greatest distance from the trend line should be an equal distance from the trend line, representing the largest margin of error.