**MACHINE LEARNING**

**ASSINGMENT - 39**

1. Which of the following methods do we use to find the best fit line for data in Linear Regression? **A) Least Square Error**

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

**A) 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?

**B) Correlation**

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?

**C) Sensitivity and Specificity**

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

**B) False**

11. Pick the feature extraction from below:

**B) Apply PCA to project high dimensional data**

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

**A) We don’t have to choose the learning rate. B) It becomes slow when number of features is very large. C) We need to iterate.**

13. Explain the term regularization?

**Ans: 1.** 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. Sometimes the machine learning model performs well with the training data but does not perform well with the test data.

Lets say we have f1,f2,f3 and these are labels, there is some data under f1,f2and f3 .We have just passed this data to our Model and Our Model sometimes process the data and gives result of 95% or 100% etc (Prediction)in training phase. This is basically in simple terms known as Overfitting.

Whenever there is overfitting problem, during testing time our model will not give accuracy as training phase of 95% and 100%. To avoid this overfitting, we perform regularization. The idea behind regularization is – it will try to apply some penalty that will reduce the weights of data in order to predict the label.

14. Which particular algorithms are used for regularization?

**Ans:** There are different types of Regularization:

L1(Lasso):

Lets say we have f1,f2,f3 and these are label, there is some data under f1,f2,f3. Lets consider f1 is contributing less as compared to f2and f3 to predict the label, then in Laaso, the f1will be considered as zero and f2,f3 will continue to predict for the label.

L2(Ridge):

Lets say we have f1,f2,f3 and these are label, there is some data under f1,f2,f3. Lets consider f1 is contributing less as compared to f2and f3 to predict the label, then in Ridge, the f1 will be given very very less importance and here f1,f2,f3 will continue to predict for the label

Elastic-Net Regression :

Elastic-Net is a regularized regression method that linearly combines the L1 and L2 penalties of the LASSO and Ridge methods respectively. The regularization parameter *(λ)* regularizes the coefficients such that if the coefficients take large values, the loss function is penalized.

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

**Ans:** In Linear Regression Equation, lets say we have X(a,b) and Y(c,d) and such we have 4or 5 points and those are plotted on the graph and a straight line is drawn such that the points are near to the straight line. But when the value is considered from the graph with the straight line the value is different. This difference in the values is known as Error.