MACHINE LEARNING

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

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

Ans. Linear regression is sensitive to outliers

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

Ans. Negative

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

Ans. Correlation

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

Ans. Low bias and high variance

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

Ans. Predictive model

7. Lasso and Ridge regression techniques belong to _____?

Ans. Regularization

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

Ans. SMOTE

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

Ans. TPR and FPR

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

Ans. False

11. Pick the feature extraction from below:

Ans. 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.

D) It does not make use of dependent variable.

Ans. A and B

13. Explain the term regularization?

Ans. Regularization is the process of adding information in order to solve an ill posed problem or to prevent overfitting. The regularization term, or penalty, imposes a cost on the optimization function to make the optimal solution unique. In machine learning, the data term corresponds to the training data and the regularization is either the choice of the model or modifications to the algorithm. It is always intended to reduce the generalization error, i.e. the error score with the trained model on the evaluation set and not the training data. One of the earliest uses of regularization is related to the method of least squares.

14. Which particular algorithms are used for regularization?

Ans. There are three main regularization techniques, namely:

a. Ridge Regression (L2 Norm)

b. Lasso (L1 Norm)

c. Dropout

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

Ans. The linear regression model contains an error term that is represented by ε. The error term is used to account for the variability in y that cannot be explained by linear relationship between x and y. If ε were not present, that would mean that knowing x would provide enough information to determine the value of y.