ASSIGNMENT – 39
MACHINE LEARNING

In Q1 to Q11, only one option is correct, choose the correct option:

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

2. Which of the following statement is true about outliers in linear regression?
Ans: A) Linear regression is sensitive to outliers

3. A line falls from left to right if a slope is _____?
Ans: B) Negative

4. Which of the following will have symmetric relation between dependent variable and independent
variable?
Ans: C) Both of them

5. Which of the following is the reason for over fitting condition?
Ans: C) Low bias and high variance

6. If output involves label then that model is called as:
Ans: B) Predictive model

7. Lasso and Ridge regression techniques belong to _____?
Ans: B) Removing outliers

8. To overcome with imbalance dataset which technique can be used?
Ans: D) SMOTE

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

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

11. Pick the feature extraction from below:
Ans: B) Apply PCA to project high dimensional data

In Q12, more than one options are correct, choose all the correct options:
12. Which of the following is true about the Normal Equation used to compute the coefficient of the Linear Regression?
Ans:
A) We don't have to choose the learning rate.
B) It becomes slow when the number of features is very large.

ASSIGNMENT – 39
MACHINE LEARNING

Q13 and Q15 are subjective answer type questions, Answer them briefly.
13. Explain the term regularization?
Regularizations are techniques used to reduce the error by fitting a function appropriately on the given training set and avoid overfitting.
machine learning, regularization is the process which regularizes or shrinks the coefficients towards zero. In simple words, regularization discourages learning a more complex or flexible model, to prevent overfitting.

14. Which particular algorithms are used for regularization?
1. Unsupervised
   A. Clustering & Dimensionality Reduction
      ● PCA
      ● K-Means
   B. Association Analysis
      ● Apriori
      ● FP Growth
   C. Hidden Markov model

2. Supervised
   A. Regression
      ● Linear
      ● Polynomial
   B. Classification
      ● KNN
      ● Tree
      ● Logistic Regression
      ● Naive-Bayes
      ● SVM
      ● Random Forest

15. Explain the term error present in linear regression equation?
Ans: An error term is a residual variable produced by a statistical or mathematical model, which is created when the model does not fully represent the actual relationship between the independent variables and the dependent variables. As a result of this incomplete relationship, the error term is the amount at which the equation may differ during empirical analysis

Error Term Use in a Formula

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:

\begin{aligned} &Y = \alpha X + \beta \rho + \epsilon \\ &\textbf{where:} \\ &\alpha, \beta = \text{Constant parameters} \\ &X, \rho = \text{Independent variables} \\ &\epsilon = \text{Error term} \\ \end{aligned}

$$Y = \alpha X + \beta \rho + \epsilon$$
where:
$$\alpha, \beta = \text{Constant parameters}$$
$$X, \rho = \text{Independent variables}$$
$$\epsilon = \text{Error term}$$