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?
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?
C) Both of them

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:
A) Descriptive model

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.
B) False

11. Pick the feature extraction from below:
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 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.

Q13 and Q15 are subjective answer type questions, Answer them briefly.

13. Explain the term regularization?
Ans.: Regularization is a technique used to reduce errors by fitting the function appropriately on the given training set and avoiding overfitting. The commonly used regularization techniques are;
1)Lasso Regularization
2)Ridge Regularization
3)Elastic Net Regularization

14. Which particular algorithms are used for regularization?
Ans.: Algorithms used for regularization are;
1) Ridge Regression or Ridge Regularization
2) Lasso(Least Absolute and Selection Operator) Regression
Ridge regression is mostly used to reduce the overfitting in the model, and it includes all the features present in the model. It reduces the complexity of the model by shrinking the coefficients. Lasso regression helps to reduce the overfitting in the model as well as feature selection

15. Explain the term error present in linear regression equation?
Ans.: Error present in linear regression also called Residual is the difference between actual value and the predicted value.
Represented as r=y-(mx+c)

$$r_i = y_i - (mx_i + b) \quad \text{(Residual for one point)}$$

$$\sum_{i=1}^{n} r_i = \sum_{i=1}^{n} (y_i - (mx_i + b)) \quad \text{(Sum of residuals)}$$

$$R(x) = \sum_{i}^{n} r_i^2 = \sum^{n} (y_i - (mx_i + b))^2 \quad \text{(Sum of squares of residuals)}$$

With the above residual calculation of R(x) will help to increase the prediction accuracy by taking steps to minimize this error.