MACHINE LEARNING (ANSWERS)

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? D) Both A and B

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? A) Regression

5. Which of the following is the reason for over fitting condition? A) High bias and high variance

6. If output involves label then that model is called as: D) All of the above

7. Lasso and Ridge regression techniques belong to _____? D) Regularization

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

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: 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.

ASSIGNMENT – 39 MACHINE LEARNING Q13 and Q15 are subjective answer type questions, Answer them briefly.

13. Explain the term regularization?

Regularization is a technique used in statistical and machine learning models to prevent overfitting, which occurs when a model is too complex and performs well on the training data but poorly on unseen data. Regularization methods add a penalty term to the objective function or loss function of a model, which helps in reducing the complexity of the model and prevents overfitting.

The following are the different types of regularization techniques commonly used in machine learning:

a) L1 Regularization (Lasso Regularization).

b)L2 Regularization (Ridge Regularization).

c)Elastic Net Regularization.

14. Which particular algorithms are used for regularization?

The following are the algorithms used for regularization:

a) Ridge Regression.

b) Lasso Regression.

c) Elastic Net Regression.

d) Logistics Regression with L1 or L2 Regularization.

e) Support Vector Machine (SVM) with Regularization.

F) Neural Networks Regularization.

15. Explain the term error present in linear regression equation

The term "error" in the context of linear regression refers to the difference between the observed values of the dependent variable and the predicted values obtained from the linear regression equation. In other words, it represents the unexplained variability in the data that cannot be accounted for by the linear relationship between the dependent variable and the independent variable(s) captured by the regression equation.

The error in a linear regression equation is also referred to as "residuals". Residuals are calculated as the differences between the observed values of the dependent variable and the corresponding predicted values obtained from the linear regression equation. Mathematically, the residual for each data point is given by:

Residual = Observed value of dependent variable - Predicted value from linear regression equation