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- Both A and B

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- Both of them

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

Ans- Low bias and low variance

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

Ans- B) Predictive modal

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- Construction bag of words from a email

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?

Ans- A) We don't have to choose the learning rate.

B) It becomes slow when number of features is very large.

13. Explain the term regularization?

Ans- Regularization is a technique used in machine learning to prevent overfitting by adding a penalty term to the model's loss function, discouraging overly complex models. It helps improve generalization performance by controlling the complexity of the model.

14. Which particular algorithms are used for regularization?

Ans- Algorithms commonly used for regularization include:

1. Ridge Regression
2. Lasso Regression
3. Elastic Net Regression
4. Logistic Regression with Ridge/Lasso/Elastic Net regularization
5. Support Vector Machines with L1/L2 regularization
6. Neural Networks with Weight Decay and Dropout regularization
7. Decision Trees with pruning techniques
8. Gradient Boosting Machines with regularization parameters.

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

Ans- In a the context of linear regression, the term "error" refers to the discrepancy between the actual observed values of the dependent variable and the values predicted by the linear regression models.

The linear regression equation can be written as:

$$ Y = \beta_0 + \beta_1X + \varepsilon $$

Where:

- Y is the dependent variable,

- X is the independent variable,

- β_0 is the intercept,

- β_1 is the slope,

- ε represents the error term.

The error term ε captures the difference between the observed values of Y and the values predicted by the linear regression model. It represents the influence of factors not included in the model, random variation, or measurement error.

The assumptions of linear regression include that the error term has a mean of zero, is independent and identically distributed (i.i.d.), and follows a normal distribution. These assumptions ensure the validity of statistical inference and the accuracy of parameter estimates derived from the linear regression model.