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?

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?

B) Correlation

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

B) Low bias and low variance

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

B) Predictive modal

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?

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?

This technique used in machine learning and statistical modeling to prevent overfitting by adding a penalty to the model's complexity.

Overfitting occurs when a model performs well on the training data but poorly on the test data.

14. Which particular algorithms are used for regularization?

To prevent overfitting and underfitting we use several models:

1. Linear Models

2. Logistic Regression

3. Neural Networks

4. Support Vector Machines

5. Decision Trees and Ensemble Methods

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

In linear regression, this term refers to the difference between the actual observed value and the predicted value of the dependent variable.

It represents the portion of the data that the model cannot explain. In statistical terms, it's often called the error term, and it accounts for the variability

in the data that is not captured by the model.