**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 B) Maximum Likelihood

C) Logarithmic Loss **D) Both A and B**

**Answer:** 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** B) linear regression is not sensitive to outliers

C) Can’t say D) none of these

**Answer:** A) Linear regression is sensitive to outliers

3. A line falls from left to right if a slope is \_\_\_\_\_\_?

A) Positive **B) Negative**

C) Zero D) Undefined

**Answer:** B) Negative

4. Which of the following will have symmetric relation between dependent variable and independent

variable?

A) Regression B) Correlation

C) Both of them **D) None of these**

**Answer**: D) None of these

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

A) High bias and high variance B) Low bias and low variance

**C) Low bias and high variance** D) none of these

**Answer**: C) Low bias and high variance

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

A) Descriptive model **B) Predictive modal**

C) Reinforcement learning D) All of the above

**Answer:** B) Predictive modal

7. Lasso and Ridge regression techniques belong to \_\_\_\_\_\_\_\_\_?

A) Cross validation B) Removing outliers

C) SMOTE **D) Regularization**

**Answer:** D) Regularization

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

A) Cross validation B) Regularization

C) Kernel **D) SMOTE**

**Answer:** 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** B) Sensitivity and precision

C) Sensitivity and Specificity D) Recall and precision

**Answer**: A) TPR and FPR

10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the

curve should be less.

1. True
2. **False**

**Answer**: False

11. Pick the feature extraction from below:

A) Construction bag of words from a email

**B) Apply PCA to project high dimensional data**

C) Removing stop words

D) Forward selection

**Answer**: 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.

**D) It does not make use of dependent variable.**

**Answer**: A & D

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

13. Explain the term regularization?

Regularization is a method employed in machine learning to avoid overfitting and enhance a model's capability to generalize. It works by introducing a penalty term into the loss function during training, which discourages the model from excessively fitting the training data or becoming overly intricate.

14. Which particular algorithms are used for regularization?

Two common types of regularization techniques are L1 regularization (Lasso) and L2 regularization (Ridge) -

**L1 regularization** adds the absolute value of the coefficients as a penalty term, encouraging sparsity and promoting feature selection. It tends to drive some coefficients to exactly zero, effectively eliminating irrelevant features from the model.

**L2 regularization** adds the square of the coefficients as a penalty term. It does not promote sparsity as strongly as L1 regularization but can help to reduce the impact of correlated features by shrinking their coefficients.

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

The error in linear regression represents the unexplained variability or noise in the data that is not captured by the linear relationship between the independent variables and the dependent variable. Minimizing this error is the objective of linear regression modeling.