MACHINE LEARNING ASSIGNMENT

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**

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**

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

A) Positive **B) Negative**

C) Zero **D) Undefined**

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**

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**

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**

7. Lasso and Ridge regression techniques belong to _____?

A) Cross validation **B) Removing outliers**

C) SMOTE **D) Regularization**

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

A) Cross validation **B) Regularization**

C) Kernel 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

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

A) True B) 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

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.

13. Explain the term regularization?

Ans: - Regularization is like putting a leash on a model to keep it from running wild. Without it, models tend to memorize data too much, which isn't good in reality. Regularization helps keep the model on track to learn simple and accurate rules. Think of it like a detective who initially remembers a lot of details but later learns to focus only on the important ones. Similarly, regularization reminds the model not to remember too many details and only learn the important rules.

14. Which particular algorithms are used for regularization?

Ans: - Some commonly used algorithms for regularization include:

- Ridge Regression: This algorithm adds a penalty term proportional to the square of the magnitudes of the coefficients (L2 regularization).

- **Lasso Regression: Similar to Ridge Regression, Lasso Regression adds a penalty term, but this one is proportional to the absolute values of the coefficients (L1 regularization).**
- **Elastic Net: This algorithm combines both L1 and L2 regularization, providing a balance between the two techniques.**
- **Support Vector Machines (SVM): SVMs use regularization parameters like 'C' to control the trade-off between maximizing the margin and minimizing the classification error.**
- **Neural Networks: Techniques like L1 or L2 regularization, dropout, and early stopping are commonly used to regularize neural networks and prevent overfitting.**

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

Ans:- In linear regression, the "error" refers to the difference between the predicted values generated by the regression equation and the actual observed values in the dataset. It represents how much our predictions deviate from the real data points.

For example, if we are using linear regression to predict house prices based on their sizes, the error would be how much our predicted prices differ from the actual sale prices of the houses. If we predict a house to be sold for $300,000 but it actually sells for $320,000, the error would be $20,000.