MACHINE LEARNING ASSIGNMENT

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

Ans. Least square error

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

Ans. 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

Ans. 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

Ans. Correlation

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

Ans. 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

Ans. Predictive model

7. Lasso and Ridge regression techniques belong to _____?

A) Cross validation B) Removing outliers

C) SMOTE D) Regularization

Ans. Regularization

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

A) Cross validation B) Regularization

C) Kernel D) SMOTE

Ans. 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

Ans. TPR and FPR

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

A) True B) False

Ans. False

11. Pick the feature extraction from below:

A) Construction bag of words from an email

B) Apply PCA to project high dimensional data

C) Removing stop words

D) Forward selection

Ans. Construction bag of words from an 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?

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.

Answers. We don't have to choose the learning rate.

It becomes slow when number of feature is very large .

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

13. Explain the term regularization?

Ans. Regularization is a technique used to prevent overfitting by adding a penalty term to the loss function. It discourages complex models with high variance that fit the noise in the data. Two common types are Lasso (which uses L1 regularization) and Ridge (which uses L2 regularization).

14. Which particular algorithms are used for regularization?

Ans. 1. Lasso Regression (L1 Regularization)-- Adds an absolute value of coefficients penalty term to the loss function.

2. Ridge Regression (L2 Regularization)--Adds a squared value of coefficients penalty term to the loss function.

3. Elastic Net--Combines Lasso and Ridge regression to balance between L1 and L2 regularization.

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

Ans. In linear regression, the error (or residual) is the difference between the actual value of the dependent variable and the predicted value by the model. This error term represents the degree of inaccuracy in the model's predictions and is minimized during model training to improve the accuracy of the predictions.