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
   Ans) A) 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) 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
   Ans) A) Positive

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) A) Regression

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) C) Low bias and low 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) B) Predictive model

7. Lasso and Ridge regression techniques belong to_____?
   A) Cross validation B) Removing outliers
   C) SMOTE D) Regularization
   Ans) D) Regularization

8. To overcome with imbalance dataset which technique can be used?
   A) Cross validation B) Regularization
   C) Kernel D) SMOTE
   Ans) 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
   Ans) A) 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) A)True

11. Pick the feature extraction from below:
    A) Construction bag of words from a email
    B) Apply PCA to project high dimensional data

MACHINE LEARNING

C) Removing stop words
D) Forward selection

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.

Ans) A) & C)

MACHINE LEARNING

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

13. Explain the term regularization?

    Ans) It is technique used to prevent overfitting, it works by adding a penalty term to the standard objective function that machine learning models aim to minimize during training. During the training, the model minimizes the total loss, striking a balance between fitting the training data well and keeping the model parameters within certain bounds.

14. Which particular algorithms are used for regularization?

    Ans) The main regularization techniques are

    Lasso Regression(L1 norm): It is also known as L1 regressions, its objective is to minimize the sum of squared differences between predicted value and the actual value. This are useful for high-dimensional datasets where many features may be irrelevant.

    Ridge Regression(L2 norm): It is also known as Tikhonov regularization, it is a technique that is used to analyze multiple regression data which is multicollinear in nature. It is not used very widely.

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

    Ans) For explaining this, we will see the main objective of linear regression is to find the best-fitting line that minimizes the sum of squared errors. Errors basically means the difference between the observed values and the values predicted by the linear regression model. Mathematically, for each data point the error is calculated as the actual value minus the predicted value. Hence the goal of the linear regression is to minimize the errors.