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- 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

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- 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

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

MACHINE LEARNING

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.
   A) True
   B) False

Answer- 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

Answer- A(Construction bag of words from a 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.

Answer- A (We don't have to choose the learning rate)

MACHINE LEARNING

B (It becomes slow when number of features is very large)

MACHINE LEARNING

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

13. Explain the term regularization?

Answer- Regularization is a technique used in machine learning to prevent overfitting by adding a penalty to the loss function. This penalty discourages the model from becoming too complex by penalizing large coefficients in the regression equation. Regularization helps improve the model's generalization by ensuring it captures the underlying pattern in the data rather than the noise.

14. Which particular algorithms are used for regularization?

Answer- The two most common regularization techniques are:

Lasso Regression (L1 Regularization): Adds a penalty equal to the absolute value of the magnitude of coefficients.

Ridge Regression (L2 Regularization): Adds a penalty equal to the square of the magnitude of coefficients.

Elastic Net: Combines both L1 and L2 regularization techniques.

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

Answer- The error in a linear regression equation, also known as the residual, is the difference between the observed value and the value predicted by the model. It represents the portion of the dependent variable that cannot be explained by the independent variables. Minimizing this error is the goal of regression analysis, as it indicates the accuracy of the model in capturing the relationship between variables.