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: 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: 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
   Ans: B) 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: 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
   Ans: B) Predictive modal

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 (Synthetic Minority Over-sampling Technique)

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: B) False

MACHINE LEARNING

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

Ans: 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.

Ans: Option A,B & D are correct

MACHINE LEARNING

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

13. Explain the term regularization?

Ans: It's a technique which used in ML to preventing the Overfitting in the models.
It involves adding a penalty term to the objective function of a model to discourage overly complex or extreme parameter values.
The penalty encourages the model to favor simpler by limiting the magnitude of coefficients.
It helps for achieving better Generalization to unseen data by reducing the model's sensitivity to fluctuations in the training data & it's improving performance on new, unseen examples.

14. Which particular algorithms are used for regularization?

Ans: These methods are applied in ML to prevent overfitting. Some popular algorithms that are used in the Regularization techniques include Ridge Regression, Lasso Regression & Elastic Net. They are modifying the basic linear regression by adding the penalty terms to the model's optimization process. Ridge Regression penalizes the squared magnitude of coefficients (L2 regularization), Lasso Regression penalizes the absolute magnitude of coefficients (L1 regularization), and Elastic Net combines both L1 and L2 penalties, offering a balanced approach between them. These techniques help in creating more robust and generalized models by controlling the complexity of the model.

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

Ans: In the linear regression, the error refers to the diffrence between the predicted values and the actual observed values of the dependent variable. It represents the deviation or the difference between where the regression line predicts the outcome for a particular data point and the actual value of the dependent variable for that point. This error, often termed residual, is the vertical distance between the observed data point and the fitted regression line. Minimizing these errors through techniques like the least squares method helps in finding the best-fitting line to the data. Beginners can understand it as the unavoidable difference between the predicted value and the actual value in a linear regression model, which we aim to minimize to create a more accurate model.