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?
   Ans: A) Least Square Error

2. Which of the following statement is true about outliers in linear regression?
   Ans: A) Linear regression is sensitive to outliers.

3. A line falls from left to right if a slope is_____?
   Ans: B) Negative

4. Which of the following will have symmetric relation between dependent variable and independent variable?
   Ans: B) Correlation

5. Which of the following is the reason for over fitting condition?
   Ans: B) Low bias and low variance

6. If output involves label then that model is called as:
   Ans: B) Predictive model.

7. Lasso and Ridge regression techniques belong to_____?
   Ans: D) Regularization.

8. To overcome with imbalance dataset which technique can be used?
   Ans: D) SMOTE.

9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses_____to make graph?
   Ans: A) TPR and FPR.

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

11. Pick the feature extraction from below:
    Ans: B) Apply PCA to project high dimensional data.

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?
    Ans: 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.

MACHINE LEARNING

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

13) Explain the term regularization?

Ans: Regularization in machine learning is a technique used to prevent overfitting and improve the generalization of a model. Overfitting occurs when a model learns to perform well on the training data but fails to generalize to new, unseen data. Regularization methods add a penalty to the models loss functions to discourage complex or extreme parameter values which can help prevent overfitting. Common types of regularization includes L1 regularization, L2 regularization, and elastic net regularization each of which applies different types of penalties to the models parameters. Regularization is an important tool for improving the performance and reliability of machine learning models.

14) Which particular algorithms are used for regularization?

Ans: There are several main regularization algorithms used in machine learning.

1) L1 Regularization (Lasso Regression): This algorithms adds a penalty term to the cost function proportional to the absolute value of the coefficients.

2) L2 Regularization (Ridge Regression): This algorithms adds a penalty term to the cost function proportional to the square of the coefficients.

3) Dropout Regularization: This algorithms randomly drops out some neurons during training to prevent overfitting.

4) Early Stopping: This algorithms stops the training process when the performance on a validation set stops improving.

5) Data Augmentation: This method involves creating new training examples by transforming the existing ones, such as rotating, scaling or flipping images.

This all algorithms can helps prevent overfitting and improve the generalization performance of machine learning models.

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

Ans: In a regression analysis, the term error represents the difference between the observed value of the dependent variable and the value predicted by the regression model. It captures the variability in the dependent variable that is not explained by the independent variables included in the model. The error term is an important component of regression analysis as it reflects the presence of unobserved factors or random variation that affect the dependent variable. It is typically assumed to follow certain statistical properties, such as being normally distributed with a mean of zero.