MACHINE LEARNING

Multiple-Choice Questions (Q1 to Q12)

1. **Which of the following methods do we use to find the best fit line for data in Linear Regression?**
   - A) Least Square Error

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

3. **A line falls from left to right if a slope is _____?**
   - B) Negative

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

5. **Which of the following is the reason for overfitting condition?**
   - C) Low bias and high variance

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

7. **Lasso and Ridge regression techniques belong to _____?**
   - D) Regularization

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

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

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

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.

Subjective Questions (Q13 to Q15)

Q13. Explain the term regularization.

Regularization is a technique used in machine learning to prevent overfitting by adding a penalty term to the model's loss function. This penalty term discourages the model from fitting the noise in the training data, thus improving its generalization to new, unseen data. Regularization helps to keep the model weights small, which can result in simpler models that are less likely to overfit.

There are two common types of regularization: L1 regularization (Lasso) and L2 regularization (Ridge). L1 regularization adds the absolute value of the coefficients as a penalty term, which can result in sparse models where some coefficients are exactly zero. L2 regularization adds the squared value of the coefficients as a penalty term, which tends to distribute the penalty more evenly across all coefficients, leading to smaller but non-zero coefficients.

Q14. Which particular algorithms are used for regularization?

The algorithms commonly used for regularization include:

- **Lasso Regression (L1 Regularization):** This algorithm adds a penalty term equal to the absolute value of the coefficients to the loss function. It can result in some coefficients being exactly zero, effectively performing feature selection.
- **Ridge Regression (L2 Regularization):** This algorithm adds a penalty term equal to the square of the coefficients to the loss function. It tends to shrink the coefficients but does not set any of them to zero.
- **Elastic Net Regression:** This algorithm combines both L1 and L2 regularization. It adds a penalty term that is a linear combination of the L1 and L2 penalties. This allows it to benefit from both Lasso and Ridge regularization.

Q15. Explain the term error present in the linear regression equation.

In linear regression, the error term (also known as the residual) represents the difference between the actual observed values and the values predicted by the model. The error term accounts for the variability in the data that cannot be explained by the linear relationship between the dependent and independent variables.

It represents the amount by which the prediction deviates from the actual value. The goal of linear regression is to minimize these errors to create a model that best fits the data.