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

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?
   C) Both of them

5. Which of the following is the reason for over fitting condition?
   D) none of these

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

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

    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?
    A) We don't have to choose the learning rate.
    B) It becomes slow when number of features is very large.
    C) It does not make use of dependent variable.

MACHINE LEARNING

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

13. **Explain the term regularization?**

ANS:

Regularization is a method to balance overfitting and underfitting a model during training. Both overfitting and underfitting are problems that ultimately cause poor predictions on new data.

14. **Which particular algorithms are used for regularization?**

ANS:

Regularization Algorithms

- **Ridge regression – Its purpose is to overcome problems such as data overfitting and multicollinearity in data. When there is considerable collinearity (the existence of near-linear connections among the independent variables) among the feature variables, a typical linear or polynomial regression model will fail. Ridge Regression adjusts the variables by a modest squared bias factor. The feature variable coefficients are pulled away from this rigidity by such a squared bias factor, providing a little bit of bias into the model but considerably lowering variation.**

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

ANS:

In a linear regression model, the error term is the difference between the model's predicted value and the actual value. It's a random variable with a constant variance and a mean of zero. The error term is made up of two parts:

- **Population error: The part of Y that is not explained by the linear combination of X and β**
- **Residuals: The sample part of Y that is not explained by X and ˆβ**