

# MACHINE LEARNING

**In Q1 to Q11, only one option is correct, choose the correct option:**

- D** 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

**A** 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

**B** 3. A line falls from left to right if a slope is \_\_\_\_\_?  
A) Positive                                        B) Negative  
C) Zero    D) Undefined

**C** 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

**C** 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

**B** 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

**D** 7. Lasso and Ridge regression techniques belong to \_\_\_\_\_?  
A) Cross validation                            B) Removing outliers  
C) SMOTE                                        D) Regularization

**D** 8. To overcome with imbalance dataset which technique can be used?  
A) Cross validation                            B) Regularization  
C) Kernel                                        D) SMOTE

**A** 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

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

**B** 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

**In Q12, more than one options are correct, choose all the correct options:**

- A** **B** 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.

## **MACHINE LEARNING**

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

13. Explain the term regularization?
14. Which particular algorithms are used for regularization?
15. Explain the term error present in linear regression equation?

**13. Regularization in machine learning is a method to prevent overfitting by penalizing complex models. It adds a regularization term to the model's objective function, discouraging overly intricate models that may fit the training data too closely. This helps improve the model's generalization to new, unseen data.**

**14. Two popular algorithms used for regularization in machine learning are:**

**Lasso (Least Absolute Shrinkage and Selection Operator):**

**Adds the absolute values of the coefficients as a penalty term. Encourages sparsity in the model, leading to some coefficients becoming exactly zero.**

**Ridge Regression:**

**Adds the squared values of the coefficients as a penalty term. Encourages smaller and more evenly distributed coefficients.**

**15. In linear regression, "error" refers to the difference between the actual data points and the values predicted by the regression model. The goal of linear regression is to minimize these errors, ensuring the model best fits the observed data.**

---