# 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: 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
   Ans: (A)

3. A line falls from left to right if a slope is_____?
   a. Positive B) Negative
   C) Zero D) Undefined
   Ans: (B)

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)

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)

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)

7. Lasso and Ridge regression techniques belong to_____?
   a. Cross validation B) Removing outliers
   C) SMOTE D) Regularization
   Ans: (D)

8. To overcome with imbalance dataset which technique can be used?
   a. Cross validation B) Regularization
   C) Kernel D) SMOTE
   Ans: (D)

# MACHINE LEARNING

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: (A)

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)

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: (B)

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

B. 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: a,b,c.

# MACHINE LEARNING

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

C. Explain the term regularization?

D. Which particular algorithms are used for regularization?

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

**Answers:**

13. Explain the term regularization?

Ans: Regularization is a set of methods for reducing overfitting in machine learning models. Typically, Regularization trades a marginal decrease in training accuracy for an increase in generalizability. Regularization encompasses a range of technique to correct for overfitting in machine learning models.

14. Which particular algorithms are used for regularization?

Ans: Algorithms used for regularization are :-

-Lasso regression (L1 regularization)

-Ridge regression (L2 regularization)

-Elastic net (L1 +L2) regularization

-Ensembling

-Neural network dropout

-Pruning decision tree-based model

-Data Argumentation

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

Ans: An error term represents the margin of error within a statistical model. It refers to the sum of deviation within the regression line, which provides an explanation for the difference between the theoretical value of the model and the actual observed results.

$$Y=\alpha X+ \beta\rho + e$$

Where,

Y = dependent variable
*α, β= constant parameter*

e = error term

X= independent variable.