# 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) Least Square Error**.

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) Linear regression is sensitive to outliers**

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

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-(C)Both of them**

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) Low bias and high variance**

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- (D) All of the above**

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

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

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) TPR and FPR**

# MACHINE LEARNING

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

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) 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) We need to iterate.
    D) It does not make use of dependent variable.

    **Ans- (A), (B) and (C)**

# MACHINE LEARNING

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

13. Explain the term regularization?

**Ans-** Regularization is a method for “constraining” or “regularizing” the size of the coefficients, thus “shrinking” them towards zero.It reduces model variance and thus minimizes overfitting.If the model is too complex, it tends to reduce variance more than it increases bias, resulting in a model that is more likely to generalize.

14. Which particular algorithms are used for regularization?

**Ans- There are three main regularization techniques, namely:**

**Ridge Regression (L2 Norm)-** Ridge regression is a method for analyzing data that suffer from multi-collinearity. Ridge regression adds a penalty ***(L2 penalty)*** to the loss function that is equivalent to the square of the magnitude of the coefficients. The regularization parameter *(λ)* regularizes the coefficients such that if the coefficients take large values, the loss function is penalized.

**Lasso (L1 Norm)Dropout-** LASSO is a regression analysis method that performs both feature selection and regularization in order to enhance the prediction accuracy of the model. LASSO regression adds a penalty ***(L1 penalty)*** to the loss function that is equivalent to the magnitude of the coefficients. In LASSO regression, the penalty has the effect of forcing some of the coefficient estimates to be **exactly equal to zero** when the regularization parameter λ is sufficiently large.

**Elastic-Net Regression-** Elastic-Net is a regularized regression method that linearly combines the L1 and L2 penalties of the LASSO and Ridge methods respectively.

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

**Ans-** In statistics, an error term is the sum of the deviations of each actual observation from a model regression line. Regression analysis is used to establish the degree of correlation between two variables, one independent and one dependent, the result of which is a line that best "fits" the actually observed values of the dependent value in relation to the independent variable or variables. Put another way, an error term is the term in a model regression equation that tallies up and accounts for the unexplained difference between the actually observed values of the independent variable and the results predicted by the model.