# 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**

**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: B) It becomes slow when number of features is very large.**

**D) It does not make use of dependent variable.**

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

**13. Explain the term regularization?**

**Ans:** Regularization is one of the basic and most important concept in the world of Machine Learning. The word regularize means to make things regular or acceptable. This is exactly why we use it for. Regularizations are techniques used to reduce the error by fitting a function appropriately on the given training set and avoid overfitting

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

**Ans:** In the regularization there are 3 types of algorithms are used. 1st is the ridge regression , 2nd on is LASSO and 3rd one is Elastic net regression . Ridge regression:- Ridge regression is a method for analyzing

data that suffer from multi-collinearity. LASSO :-LASSO is a regression analysis method that performs both feature selection and regularization in order to enhance the prediction accuracy of the model. Elastic-Net:-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:** An error term is a residual variable produced by a statistical or mathematical model, which is created when the model does not fully represent the actual relationship between the independent variables and the dependent variables. As a result of this incomplete relationship, the error term is the amount at which the equation may differ during empirical analysis. The error term is also known as the residual, disturbance, or remainder term, and is variously represented in models by the letters e, ε, or u. An error term appears in a statistical model, like a regression model, to indicate the uncertainty in the model. The error term is a residual variable that accounts for a lack of perfect goodness of fit. Heteroskedastic refers to a condition in which the variance of the residual term, or error term, in a regression model varies widely. An error term represents the margin of error within a statistical model; it refers to the sum of the deviations within the regression line, which provides an explanation for the difference between the theoretical value of the model and the actual observed results. The regression line is used as a point of analysis when attempting to determine the correlation between one independent variable and one dependent variable.