# MACHINE LEARNING

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: B) Correlation**

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: A) High 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: 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 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; B) False**

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**

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) We don't have to choose the learning rate**

13. Explain the term regularization?

Ans: Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting. Using Regularization, we can fit our machine learning model appropriately on a given test set and hence reduce the errors in it.

.

14. Which particular algorithms are used for regularization?

There are two main types of regularization techniques:

1.Ridge Regularization ：Also known as Ridge Regression, it modifies the over-fitted or under fitted models by adding the penalty equivalent to the sum of the squares of the magnitude of coefficients.

This means that the mathematical function representing our machine learning model is minimized and coefficients are calculated. The magnitude of coefficients is squared and added. Ridge Regression performs regularization by shrinking the coefficients present. The function depicted below shows the cost function of ridge regression :

2 Lasso Regularization：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.

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

Ans: The error term in a regression equation represents the effect of the variables that were omitted from the equation. By using mean square error method we can calculate the error in the model.