MACHINE LEARNING

1. Which of the following methods do we use to find the best fit line for data in Liner Regression?
   A) Least Square Error
   B) Maximum Likelihood
   C) Logarithmic Loss
   D) Both A and B

Answer for the above is – "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

Answer for the above is - "A"

3. A line falls from left to right if a slope is _____?
   A) Positive
   B) Negative

   C) Zero
   D) Undefined

Answer for the above is- "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

Answer for the above is- "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 the above

Answer for the above is- "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

Answer for the above is- "B"

7. Lasso and Ridge regression techniques belong to _____?
   A) Cross validation
   B) Removing outliers

   C) SMOTE
   D) Regularization

Answer for the above is- "D"

8. To overcome with imbalance dataset which technique can be based?
   A) Cross Validation
   B) Regularization

   C) Kernel
   D) SMOTE

Answer for the above is-**"D"**

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

Answer for the above is-"A"

10. In AUCROC curve for the better model area under the curve should be less.
    A) True   B) False

Answer for the above is- "A"

11. Pick the feature extraction from the below:
    A) Construction bag of words from a email.
    B) Apply PCA to project high dimensional data.
    C) Removing stop words
    D) Forward selection
12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?

We don't have **Answer for the above is- "D"**

    A) 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.

Answer for the above is – "A,B,C"

13. Explain the term regularization?

Regularization refers to the techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting. Choosing overly complex models for the training data points can often lead to overfitting. On the other hand, a simpler model leads to underfitting the data. Hence choosing just the right amount of complexity in the model is critical. Since the complexity of the model cannot be directly inferred from the available training data, it is often impossible to stumble upon the right model complexity for training.

Based on the approach used to overcome overfitting, we can classify the regularization techniques into three categories. Each regularization method is marked as a strong, medium, and weak based on how effective the approach is in addressing the issue of overfitting.

14. Which particular algorithms are used for regularization?

There are three different types of algorithms used for regularization.

Ridge Regression

LASSO (Least Absolute Shrinkage and Selection Operator) Regression

Elastic-Net Regression

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

The error term of a regression equation represents all of the variation in the dependent variable not explained by the weighted independent variables.

A regression equation is the formula for a straight line — in this case, the best-fit line through a scatterplot of data. If there were no error, all the data points would be located on the regression line; to the extent they are not represents error; this is what the error term summarizes.