MACHINE LEARNING -1

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: - 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: - 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: - B) Predictive modal

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 an email

B) Apply PCA to project high dimensional data

C) Removing stop words

D) Forward selection In

Ans: - B) Apply PCA to project high dimensional data

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) We don't have to choose the learning rate. &

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

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

13) Explain the term regularization?

Regularization is one of the most important concepts of machine learning. It is a technique to prevent the model from overfitting by adding extra information to it.

Regularization is one of the most important concepts of machine learning. It is a technique to prevent the model from overfitting by adding extra information to it.

One of the ways of avoiding overfitting is using cross validation, that helps in estimating the error over test set, and in deciding what parameters work best for your model.

14) Which particular algorithms are used for regularization?

Ans: -

Techniques of Regularization

Mainly, there are two types of regularization techniques, which are given below:

1. Ridge Regression
2. Lasso Regression

1. Ridge Regression: -

Ridge regression is one of the types of linear regression in which we introduce a small amount of bias, known as Ridge regression penalty so that we can get better long-term predictions.

2. Lasso Regression: -

Lasso regression is another variant of the regularization technique used to reduce the complexity of the model. It stands for Least Absolute and Selection Operator.

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

A Linear Regression model's main aim is to find the best fit linear line and the optimal values of intercept and coefficients such that the error is minimized. Error is the difference between the actual value and Predicted value and the goal is to reduce this difference. Within a linear regression model tracking a stock's price over time, the error term is the difference between the expected price at a particular time and the price that was actually observed.