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:

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 model

C) Reinforcement learning

D) All of the above

Ans:

B) Predictive model

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

10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.

A) True

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

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:

D) It does not make use of dependent variable.

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

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

13. Explain the term regularization?

Regularization is a technique used in Machine learning to prevent overfitting and improve the model.

Two different types of regularization techniques: L1 regularization(Lasso) and L2 regularization(Ridge)

It is useful when dealing with high dimensional datasets or number of features is close to or exceeds the number of observations. Also helps to balance the trade-off between fitting and the training data. Maintain good performance on unseen data.

14. Which particular algorithms are used for regularization?

Regularization is a general concept/technique that can be used in linear and non-linear models. It is used to prevent the overfitting.

a. Linear Regression with Lasso(L1) and Ridge(L2)
b. Logistic regression with Lasso(L1) and Ridge(L2)
c. Support Vector Machines (SVM)
d. Decision Trees and Random Forests

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

The "error" term refers the difference between actual values of y and values predicted by the linear model.

The differences represent the discrepancies or inaccuracy in the model predictions and they are quantified as "Residuals"

Linear equation is:

$Y=b0+b1.x1+b2.x2+...+E$

"E" – represents the error term