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 overfitting 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 Bais 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 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 imbalanced 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 a graph.

A) TPR and FPR

B) Sensitivity and precision

C) Sensitivity and Specificity

D) Recall and precision

Ans: A) TPR and FPR

10. In the 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

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 the 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 a learning rate. B) It becomes shown when a number of features is very large. C) We need to iterate

13. Explain the term regularization.

Ans: Regularization

Regularization methods is very important methods in ML . Ridge and lasso regulation can introduce penalty terms to the regression coefficients helping to reduce the impact of collinearity.

It is a way to prevent Overfitting. and thus improve the performance of the model by reducing the complexity of the final model.

In regularization, we do normally keep the same number of features. But reduces the magnitude of the coefficients.

The main objective of this is to scale down the coefficient Value.

14. Which particular algorithms are used for regularization?

As above mention regulation two types of method

1 . Ridge Regression

2 . Lasso regression

In the Ridge regression magnitude of the coefficient is almost zero, but in Lasso magnitude of the coefficient is exactly zero.

We can find the appropriate value alpha value by doing cross-validation.

Lasso is used for feature selection

Ridge is also abbreviated as L2 Regularization, and Lasso is L1 Regularization.

Formula for Ridge = Loss + alpha ||W||2

Formula for Lasso = Loss + alpha ||W||

Where

Loss = cost function (i.e “Difference” between predict and actual value)

W = Slope

Alpha = consta

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

Ans : In Linear regression equation term error also abbreviated as coat functions and its difference between observed and actual values of the dependent variable and predicted variable.

Equation of the linear regression is

Y = a + B x + e

` Error or residue or cost function

Independent variable

Coefficient of slope

Intercept

Dependant Variable

Term error it could be a random error, or measurement error the goal of the linear regression minimizes the error by the sum of squares of error or the sum of square residual this minimizes the difference between actual error and predicted error.