MACHINE LEARNING

1. Which of the following methods do we use to find the best fit line for data in Linear Regression?

Ans : A. Least square error

2. Which of the following statement is true about outliers in linear regression?

Ans: A. Linear regression is sensitive to outliers

3. A line falls from left to right if a slope is _____?

Ans: B. Negative

4. Which of the following will have symmetric relation between dependent variable and independent variable?

Ans: B. Correlation

5. Which of the following is the reason for over fitting condition?

Ans: C. Low bias and high variance

6. If output involves label then that model is called as:

Ans: B. Predictive modal

7. Lasso and Ridge regression techniques belong to _____?

Ans: D. Regularization

8. To overcome with imbalance dataset which technique can be used?

Ans: D. SMOTE

9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph.

Ans: A.TPR and FPR

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

Ans: B. False

11. Pick the feature extraction from below:

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?

Ans: 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.

13. Explain the term regularization.

Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent over fitting or under fitting. 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.
2) Lasso Regularization.

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.

$$\text{Cost function} = \text{Loss} + \lambda \times \sum \|w\|^2$$

In the cost function, the penalty term is represented by Lambda λ. By changing the values of the penalty function, we are controlling the penalty term. The higher the penalty, it reduces the magnitude of coefficients. It shrinks the parameters. Therefore, it is used to prevent multi co-linearity, and it reduces the model complexity by coefficient shrinkage.

2) **Lasso Regularization**. It modifies the over-fitted or under-fitted models by adding the penalty equivalent to the sum of the absolute values of coefficients. Lasso regression also performs coefficient minimization, but instead of squaring the magnitudes of the coefficients, it takes the true values of coefficients.

$$\text{Cost function} = \text{Loss} + \lambda \times \sum \|w\|$$

This means that the coefficient sum can also be 0, because of the presence of negative coefficients. the cost function for Lasso regression. We can control the coefficient values by controlling the penalty terms, just like we did in Ridge Regression.

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

An error term represents the margin of error within a statistical model; it refers to the sum of the deviations within the regression line, which provides an explanation for the difference between the theoretical value of the model and the actual observed results. The regression line is used as a point of analysis when attempting to determine the correlation between one independent variable and one dependent variable.

An error term essentially means that the model is not completely accurate and results in differing results during real-world applications. For example, assume there is a multiple linear regression function that takes the following form:

$$Y=\alpha X+\beta\rho+\epsilon$$

where:

α, β=Constant parameters

X,ρ=Independent variables

ϵ=Error term

When the actual Y differs from the expected or predicted Y in the model during an empirical test, then the error term does not equal 0, which means there are other factors that influence Y.

:

MACHINE LEARNING

.