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: C) Both of them

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 model

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: C) Sensitivity and Specificity

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

Ans: B) False

11. **N/A**

12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?

Ans: A), B)

Ans 13/14: Regularization-This is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoiding overfitting**.** Using Regularization, we can fit our machine learning model appropriately on a given test set and hence reduce the errors in it.

Regularization Techniques- There are two main types of regularization techniques: Ridge Regularization (L2) and Lasso Regularization (L1).

*Ridge Regularization- Ridge regression is one of the types of linear regression in which a small amount of bias is introduced so that we can get better long-term predictions. This technique is used to reduce the complexity of the model. It is also called L2 regularization.

In this technique, the cost function is altered by adding the penalty term to it. The amount of bias added to the model is called Ridge Regression penalty. We can calculate it by multiplying with the lambda to the squared weight of each individual feature.

If the values of λ tend to zero, the equation becomes the cost function of the linear regression model. Hence, for the minimum value of λ, the model will resemble the linear regression model. A general linear or polynomial regression will fail if there is high collinearity between the independent variables, so to solve such problems, Ridge regression can be used. It helps to solve the problems if we have more parameters than samples.

*Lasso Regularization- Lasso regression is another regularization technique to reduce the complexity of the model. It stands for Least Absolute and Selection Operator. It is similar to the Ridge Regression except that the penalty term contains only the absolute weights instead of a square of weights. Since it takes absolute values, hence, it can shrink the slope to 0, whereas Ridge Regression can only shrink it near to 0. It is also called as L1 regularization.

Some of the features in this technique are completely neglected for model evaluation. Hence, the Lasso regression can help us to reduce the overfitting in the model as well as the feature selection.

Difference between Ridge Regression and Lasso Regression:-

Ridge regression is mostly used to reduce overfitting in the model, and it includes all the features present in the model. It reduces the complexity of the model by shrinking the coefficients.

Lasso regression helps to reduce the overfitting in the model as well as feature selection.

Ans 15: Error Term- An error term is a residual variable produced by a statistical or mathematical model, which is created when the model does not fully represent the actual relationship between the independent variables and the dependent variables. As a result of this incomplete relationship, the error term is the amount at which the equation may differ during analysis. The error term is a residual variable that accounts for a lack of perfect goodness of fit.

The error term is also known as the residual, disturbance, or remainder term, and is variously represented in models by the letters e, ε, or u.

$Y = a + b*X + e$

Where, a is the intercept, b is the slope of the line, and e is the error term. The equation above is used to predict the value of the target variable based on the given predictor variable(s).

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.