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

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

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

A) Positive B) Negative C) Zero D) Undefined

Ans) A

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

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

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

7. Lasso and Ridge regression techniques belong to _____?

A) Cross validation B) Removing outliers C) SMOTE D) Regularization

Ans)D

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

A) Cross validation B) Regularization C) Kernel D) SMOTE

Ans)D

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

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

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

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)B and C

13. Explain the term regularization?

Ans) It is one of the most important concepts of machine learning. This technique prevents the model from overfitting by adding **extra information** to it.

It is a form of regression that shrinks the coefficient estimates towards zero. In other words, this technique forces us not to learn a more complex or flexible model, to avoid the problem of overfitting.

For regression problems, **the increase in flexibility of a model is represented by an increase in its coefficients**, which are calculated from the regression line.

In simple words, **"In the Regularization technique, we reduce the magnitude of the independent variables by keeping the same number of variables".** It maintains accuracy as well as a generalization of the model.

14. Which particular algorithms are used for regularization?

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

- Ridge Regression
- Lasso Regression
- **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.

In Statistics, it is known as the **L-2 norm**.

In this technique, the cost function is altered by adding the penalty term (shrinkage term), which multiplies the lambda with the squared weight of each individual feature.

Usage of Ridge Regression:

- When we have the independent variables which are having high collinearity (problem of multicollinearity) between them, at that time general linear or polynomial regression will fail so to solve such problems, Ridge regression can be used.
- If we have more parameters than the samples, then Ridge regression helps to solve the problems.

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**.It is similar to the Ridge Regression except that the penalty term includes the absolute weights instead of a square of weights.

Key Differences between Ridge and Lasso Regression

Ridge regression helps us to reduce only the overfitting in the model while keeping all the features present in the model. It reduces the complexity of the model by shrinking the coefficients whereas Lasso regression helps in reducing the problem of overfitting in the model as well as automatic feature selection.

Lasso Regression tends to make coefficients to absolute zero whereas Ridge regression never sets the value of coefficient to absolute zero.

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

Ans) 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