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 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

Ans:- A) TPR and FPR

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) 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

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

D) It does not make use of dependent variable.

13. Explain the term regularization?

Ans:- Regularization is a technique used in machine learning to prevent overfitting and improve the generalization of a model. Overfitting occurs when a model learns to memorize the training data rather than capturing the underlying patterns, leading to poor performance on unseen data. Regularization introduces a penalty term to the model's objective function, which discourages overly complex models by imposing constraints on their parameters. The two most common types of regularization are L1 regularization (Lasso) and L2 regularization (Ridge).

14. Which particular algorithms are used for regularization?

Ans:- Regularization techniques can be applied to various machine learning algorithms to prevent overfitting and improve generalization. Some of the most common algorithms that incorporate regularization include:

Linear Regression: Regularization can be applied to linear regression models using techniques such as Ridge Regression (L2 regularization) and Lasso Regression (L1 regularization).

Logistic Regression: Similar to linear regression, logistic regression models can also benefit from regularization techniques like Ridge and Lasso.

Support Vector Machines (SVM): SVM models can be regularized using techniques such as the soft-margin SVM, where a regularization parameter (C) controls the trade-off between maximizing the margin and minimizing the classification error.

Neural Networks: Regularization techniques are crucial for neural networks to prevent overfitting, especially in deep learning. Common regularization methods for neural networks include L1 and L2 regularization, dropout, and early stopping.

Decision Trees: While decision trees themselves do not inherently require regularization, ensemble methods like Random Forests and Gradient Boosting Machines (GBMs) can be regularized using techniques like limiting tree depth, adjusting the number of trees, or using shrinkage.

K-Nearest Neighbors (KNN): While KNN is not typically associated with regularization in the same way as parametric models, techniques such as feature scaling and dimensionality reduction can indirectly help to regularize KNN models.

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

Ans:- In a linear regression model, the goal is to minimize these errors by finding the best-fitting line that describes the relationship between the independent variables and the dependent variable. The error term captures the part of the dependent variable that the model cannot explain using the independent variables included in the model. It reflects the influence of other factors that affect the dependent variable but are not accounted for in the model. These factors contributing to the error term could include measurement errors, unobserved variables, random variability, or inherent noise in the data. No model can perfectly capture all the complexities of real-world data, so there will always be some residual error left over after fitting the model. Understanding and analyzing the errors in a linear regression model is crucial for assessing the model's performance, identifying areas for improvement, and making reliable predictions. Common techniques for evaluating errors include calculating metrics such as mean squared error, mean absolute error, or root mean squared error, among others. Additionally, diagnostic plots, such as residual plots, can provide insights into the distribution and patterns of the errors, helping to validate the assumptions of the linear regression model.