Name: Arup Singha

Project: Batch DS2304

MACHINE LEARNING

In Q1 to Q11, only one option is correct, choose the correct option:

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: D) Both A and B

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: D) None of these

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 modal C) Reinforcement learning D) All of the above

Ans: B) Predictive modal

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

In Q12, more than one options are correct, choose all the correct options:

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 and B

Q13 and Q15 are subjective answer type questions, Answer them briefly.

13.Explain the term regularization?

Ans: Regularization refers to techniques used to calibrate machine learning models to minimize the adjusted loss function and avoid overfitting or underfitting. Overfitting occurs when a model performs well on the training data but fails to generalize well to unseen data.

Regularization achieves this by adding a penalty term to the loss function during the training process. This penalty term discourages complex and overly flexible models, favoring simpler models that are less prone to overfitting.

14. Which particular algorithms are used for regularization?

Ans: The algorithms used for regularization are:

1. Linear Regression: Regularized linear regression models, such as Ridge regression and Lasso regression, are often used to add regularization to linear regression models.
2. Logistic Regression: Regularized logistic regression, similar to regularized linear regression, applies regularization techniques to logistic regression models to prevent overfitting.
3. Support Vector Machines (SVM): SVM algorithms can be regularized using techniques like L1 regularization (Lasso SVM) or L2 regularization (Ridge SVM) to control the complexity of the decision boundary.
4. Neural Networks: Regularization techniques, such as L1 or L2 regularization (weight decay), dropout, or early stopping, can be applied to neural networks to prevent overfitting and improve generalization.

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

Ans: In linear regression, the term "error" refers to the difference between the predicted value and the actual value of the dependent variable. The goal of linear regression is to minimize this error by finding the best-fitting line that represents the relationship between the independent variables (also known as predictors or features) and the dependent variable.

The error term consists of various factors, including measurement errors, omitted variables, and random noise. These factors introduce uncertainty and imperfections into the relationship between the independent and dependent variables, resulting in a deviation between the predicted and actual values.