ASSIGNMENT-3

Annwesha Das Mohapatra-DS2312

MACHINE LEARNING ASSIGNMENT

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

1A. 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

2A. 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

3A. 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

4A. 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

5A. 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

6A. B) Predictive modal

7. Lasso and Ridge regression techniques belong to _____? A) Cross validation B) Removing outliers C) SMOTE D) Regularization

7A. D) Regularization

8. To overcome with imbalance dataset which technique can be used? A) Cross validation B) Regularization C) Kernel D) SMOTE

8A. 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

9A. 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

10A. 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

11A. 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.

12A. B) It becomes slow when number of features is very large.

C) We need to iterate.

13. Explain the term regularization?

13A. Regularization is a technique used to reduce errors by fitting the function appropriately on the given training set and avoiding overfitting. it maintains accuracy as well as a generalization of the model.

14. Which particular algorithms are used for regularization?

14A. The commonly used algorithms in regularization are

Lasso (L1 regularization) -- A regression model which uses the L1 Regularization technique is called LASSO (Least Absolute Shrinkage and Selection Operator) regression. Lasso Regression adds the "absolute value of magnitude" of the coefficient as a penalty term to the loss function(L).

Ridge (L2 regularization) regression -- A regression model that uses the L2 regularization technique is called Ridge regression. Ridge regression adds the "squared magnitude" of the coefficient as a penalty term to the loss function(L).

These techniques add a penalty term to the linear regression cost function, influencing the optimization process to favor simpler models.

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

15A. 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. In simple terms, the error refers to the difference between the predicted values by the model and the actual values in the dataset. This difference is often measured using a metric like Mean Squared Error (MSE) or Mean Absolute Error (MAE). The goal of linear regression is to minimize this error, finding the best-fit line that minimizes the difference between predicted and actual values.

PYTHON – WORKSHEET 1

1A.C)%

2A.B)0

3A.C)24

4A.A)2

5A.D)6

6A. C) the finally block will be executed no matter if the try block raises an error or not.

7A.A) It is used to raise an exception

8A.C) in defining a generator

9A.A) _abc

C) abc2

10A.A) yield

B) raise