ASSIGNMENT – 39

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

2. Which of the following statement is true about outliers in linear regression?

A) Linear regression is sensitive to outliers

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

B) Negative

4. Which of the following will have symmetric relation between dependent variable and independent variable?

B) Correlation

5. Which of the following is the reason for over fitting condition?

C) Low bias and high variance

6. If output involves label then that model is called as:

B) Predictive modal

7. Lasso and Ridge regression techniques belong to _____?

D) Regularization

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

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

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

B) False

11. Pick the feature extraction from below:

Construction bag of words from an email, Apply PCA to project high dimensional data and Removing stop words

D) All 3 above

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.

ASSIGNMENT – 39

MACHINE LEARNING

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

13. Explain the term regularization?

Regularization is a technique used to reduce the error by fitting a function appropriately on the given training set and avoid overfitting. The word regularize means to make things regular or acceptable.

A good model does not give more weight to a particular feature. The weights are evenly distributed. This can be achieved by doing regularization

14. Which particular algorithms are used for regularization?

- L1 Regularization or Lasso Regularization (Least Absolute Shrinkage and Selection Operator)
- L2 Regularization or Ridge Regularization.
- Elastic net regression.
- Dropout
- Early stopping
- Weight decay

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

An error term appears in a statistical model, like a regression model, to indicate the uncertainty in the model.

The error term is a residual variable that accounts for a lack of perfect goodness of fit

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.

Linear regression most often uses mean-square error (MSE) to calculate the error of the model. MSE is calculated by:

- measuring the distance of the observed y from the predicted y at each value of x
- squaring each of these distances
- calculating the mean of each of the squared distances

Linear regression fits a line to the data by finding the regression coefficient that results in the smallest MSE