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

Reason:-The maximizing Likehood function is equivalent to minimizing the squared differences(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

Reason:- Linear Regression is sensitive to outliers because our model will try to cater the outliers which will end up messing the best fit line.

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

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

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

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

Reason:-Low Bias means that model is not is not biased towards any particular feature and is able to understand training data high variance means prediction accuracy of training and testing data varies too much which will lead to over fitting. So correct answer is **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

7. Lasso and Ridge regression techniques belong to _____?

A) Cross validation B) Removing outliers

C) SMOTE D) Regularization

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

A) Cross validation B) Regularization

C) Kernel 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

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

A) True 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

Reason :- word frequency are used as features for text classification 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.

ASSIGNMENT _ 39

MACHINE LEARNING

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

13. Explain the term regularization?

Answer:- regularization is a technique used to avoid overfitting and underfitting by doing regularization the model generalizes the patterns learned throughout training phase to unseen testing data.Regularisation works by assigning penalty to cost function this penalty term discourages high bias to any feature i.e no feature will get preference.

14. Which particular algorithms are used for regularization?

Answer:-L1 (lasso) and L2 (Ridge)

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

Answer:- In the context of linear regression, the term "error" refers to the difference between the actual observed values of the dependent variable and the values predicted by the linear regression model. It represents the discrepancy between what the model predicts and the true outcomes in the dataset.