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- 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 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) We don't have to choose the learning rate.
b)It becomes slow when number of features is very large.

FLIP ROBO

ASSIGNMENT – 39

MACHINE LEARNING

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

13. Explain the term regularization?

Ans- When we use regression models to train some data, there is a good chance that the model will overfit the given training data set. Regularization helps sort this overfitting problem by restricting the degrees of freedom of a given equation i.e., simply reducing the number of degrees of a polynomial function by reducing their corresponding weights

14- Which particular algorithms are used for regularization?

Ans - Different types of regularizations in regression:

- LASSO
- RIDGE
- Elastic Net

➔ **LASSO (Least Absolute Shrinkage and Selection Operator) Regression (L1 Form)-**

LASSO regression penalizes the model based on the sum of magnitude of the coefficients. The regularization term is given by
regularization= $\lambda * \Sigma|| \beta j$
Where, λ is the shrinkage factor.

➔ **Ridge Regression (L2 Form)-**

Ridge regression penalizes the model based on the sum of squares of magnitude of the coefficients. The regularization term is given by
regularization= $\lambda * \Sigma|| \beta 2$
Where, λ is the shrinkage factor.

➔ **Elastic Net-**

Elastic net linear regression uses the penalties from both the lasso and ridge techniques to regularize regression models. The technique combines both the lasso and ridge regression methods by learning from their shortcomings to improve the regularization of statistical models

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

Ans- In linear regression, the "error" refers to the discrepancy between the actual observed values of the dependent variable and the values predicted by the linear regression equation. These errors quantify how well the model captures the real-world data points and represent the unexplained variability in the data. The objective of linear regression is to find the regression coefficients that minimize these errors, allowing the model to approximate the underlying relationship between the independent and dependent variables as closely as possible.