Assignment -1

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?

   ANSWER: - A) Least Square Error

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

   ANSWER: - A) Linear regression is sensitive to outliers

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

   ANSWER: - B) Negative

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

   Answer: - C) Both of them

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

   ANSWER: - C) Low bias and high variance

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

   ANSWER: - B) Predictive modal

7. Lasso and Ridge regression techniques belong to _____?

   ANSWER: - D) Regularization

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

   ANSWER: - D) SMOTE

9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses _____ to make graph?

   ANSWER: - D) Recall and precision

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

    ANSWER: - A) True

11. Pick the feature extraction from below:

    ANSWER:-

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?

ANSWER: - A) We don't have to choose the learning rate.

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

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

13. Explain the term regularization?

Answer: - Regularization can be defined as the technique which is used to reduce the errors by fitting the function in a given training set and also to avoid overfitting appropriately. There are basically 2 algorithm to avoid under or over fitting that is L1 regularization and L2 regularization also known as LASSO regression and Ridge regression respectively.

14. Which particular algorithms are used for regularization?

Answer: - There are basically two particular algorithm which are used for regularization which are

1. L1 Regularization
2. L2 Regularization

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

$$\sum_{i=1}^{n}(Y_i - \sum_{j=1}^{p} X_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

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

$$\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

The key difference between these techniques is that Lasso shrinks the less important feature's coefficient to zero thus, removing some feature altogether. So, this works well for feature selection in case we have a huge number of features.

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

Answer: - An error term is a residual variable produced by a statistical or mathematical model, which is created when the model does not fully represent the actual relationship between the independent variables and the dependent variables. This can be due to number of things not done while working with data like cleaning of data, multi-collinearity, auto correlation.

These error can be reduced by using the process of MSE or Mean Squared Error method.