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 model

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

MACHINE LEARNING

B) Apply PCA to project high dimensional data
C) Removing stop words
D) Forward selection

Ans-A) Construction bag of words from a email

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.

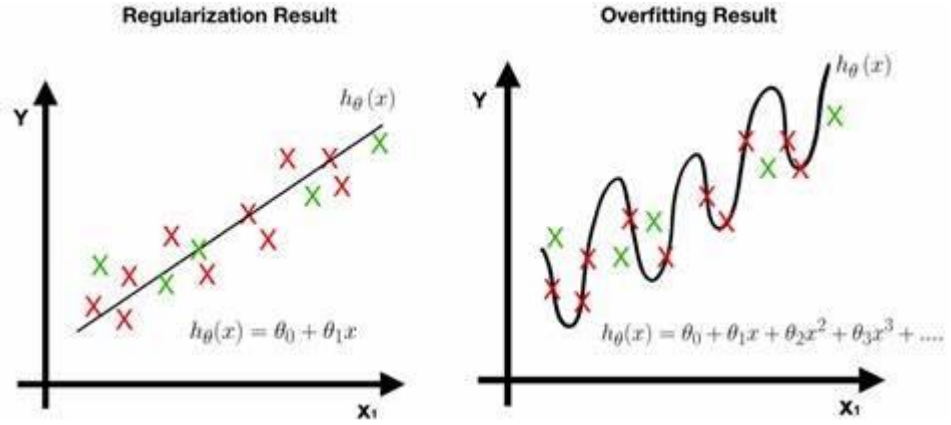
Ans-A,B,C) (A)We don't have to choose the learning rate.,,(B)It becomes slow when number of feature is vary large.,,(C)We need to iterate

MACHINE LEARNING

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

13. Explain the term regularization?

Ans->



Regularization is a technique used in machine learning to prevent overfitting of a model by adding a penalty term to the loss function. The penalty term discourages the model from fitting the training data too closely and instead encourages it to find a simpler and more generalizable solution.

In linear regression, for example, L1 and L2 regularization techniques are commonly used. L1 regularization, also known as Lasso regularization, adds a penalty term proportional to the absolute values of the coefficients. This leads to sparse solutions where many of the coefficients are set to zero. L2 regularization, also known as Ridge regularization, adds a penalty term proportional to the squared values of the coefficients. This leads to solutions where all the coefficients are shrunk towards zero but none of them are exactly zero.

Regularization can also be applied to other machine learning models, such as neural networks, support vector machines, and decision trees. The choice of regularization technique and the amount of regularization to apply is typically determined through cross-validation, where the model is trained on different subsets of the data and the performance is evaluated on a holdout set. The goal is to find the optimal balance between model complexity and generalization performance.

14. Which particular algorithms are used for regularization?

Ans-> Regularization can be applied to various machine learning algorithms. Some of the commonly used algorithms that can be regularized are:

- Linear Regression: Lasso Regression, Ridge Regression, Elastic Net Regression
- Logistic Regression: L1 Regularization, L2 Regularization
- Support Vector Machines: L1 Regularization, L2 Regularization
- Neural Networks: Dropout Regularization, L1 Regularization, L2 Regularization
- Decision Trees: Cost-Complexity Pruning
- k-Nearest Neighbors: Distance Weighting

Note that these are not the only algorithms that can be regularized, and there are many other algorithms where regularization can be applied. The choice of regularization technique and the amount of regularization to apply depends on the specific problem and the performance trade-offs that the model needs to make.

MACHINE LEARNING

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

Ans->In the context of linear regression, the term "error" refers to the difference between the predicted values of the dependent variable (y) using the linear equation and the actual values of the dependent variable.

In other words, the linear regression model tries to estimate the relationship between the independent variables (x) and the dependent variable (y) by fitting a line to the data. However, this line will not perfectly fit all the data points, and there will always be some degree of error or residual between the predicted values and the actual values.

Mathematically, the error term can be represented as follows:

error = y - y_predicted

where y is the actual value of the dependent variable, and y_predicted is the predicted value of the dependent variable based on the linear regression equation.

The goal of linear regression is to minimize the sum of squared errors (SSE) between the predicted values and the actual values. This is typically done using a technique called least squares regression, which involves finding the values of the coefficients that minimize the SSE. By minimizing the error, the linear regression model can accurately estimate the relationship between the independent variables and the dependent variable, and make predictions on new data points.