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?
   D) Both A and B
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?
   A) Regression
5. Which of the following is the reason for over fitting condition?
   A) High 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:
    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) It becomes slow when number of features is very large.
    B) We need to iterate.

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 errors by fitting the function appropriately on the given training set and avoid overfitting and prevents the loss of important data due to underfitting.

14. Which particular algorithms are used for regularization?

    2 algorithms are used for regularization

    1) L1 or Lasso – It is a a procedure that shrinks the co-efficient towards zero. L1 regularization method by modifying the RSS by adding the penalty or shrinkage quantity i.e Alpha equivalent to the value or sum of the coefficients. when Alpha equals zero, you will get the same linear regression equation's coefficient. When Alpha is equaled to infinity, the lasso regression coefficient will automatically be equal to zero. When Alpha is less than infinite, the lasso regression will run automatically between 0 and 1.

    2) L2 or Ridge- It modifies the RSS by adding the shrinkage quantity or penalty to the estimates' square, and they will become changed with the loss function. when the Alpha equals zero, then the penalty term has no effect, and you will get the same coefficients as linear regression. When Alpha (α) is infinite, it means the ridge regression coefficient will become zero because the modified loss function will not depend on core loss and at the same time minimize the coefficient square and thus taken the zero parameters.

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

    Error is the difference between the expected value at a particular time and the value that was actually observed.

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

    MSE Calculation

    Measuring the distance of the observed y-values from the predicted y-values at each value of x;

    - Squaring each of these distances;
    - Calculating the mean of each of the squared distances.