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

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

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

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

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.

MACHINE LEARNING

13. Explain the term regularization?

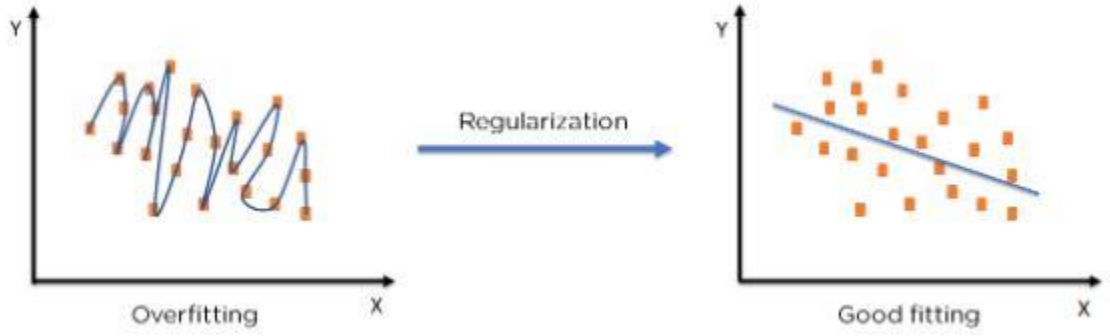
Ans.- Regularization is a method for constraining or regularizing the size of the coefficients, thus shrinking them towards zero. It reduces model variance and thus minimizes overfitting.

It reduces model variance and thus minimizes overfitting.

If the model is too complex, it tends to reduce variance more than it increases bias, resulting in a model that is more likely to generalize.

14.Which particular algorithms are used for regularization?

Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting. Using Regularization, machine learning model can fit appropriately on a given test set and hence reduce the errors in it.



Regularization on an over-fitted model

Ridge Regularization :

Also known as Ridge Regression, it modifies the over-fitted or under fitted models by adding the penalty equivalent to the sum of the squares of the magnitude of coefficients.

Lasso Regression

It modifies the over-fitted or under-fitted models by adding the penalty equivalent to the sum of the absolute values of coefficients.

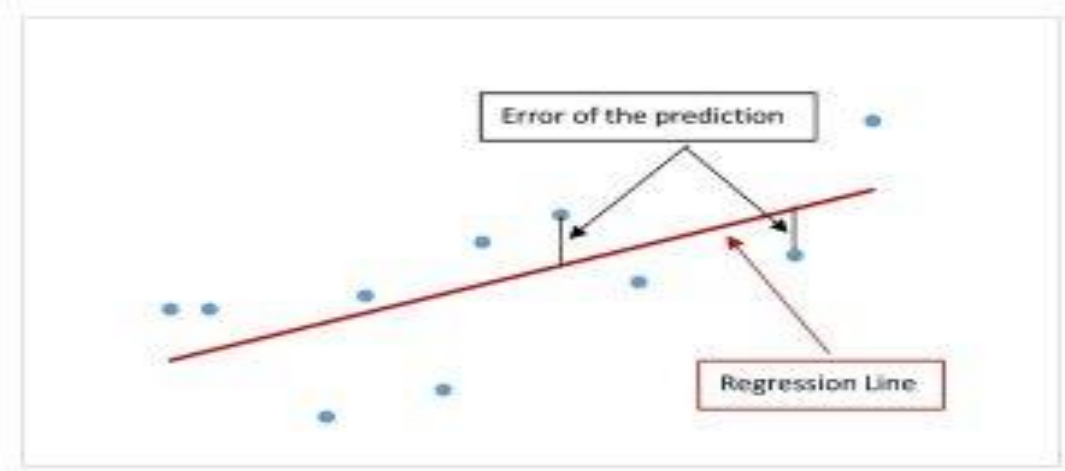
14. Explain the term error present in linear regression equation?

Ans.- The difference between the actual value of the dependent variable y (in the sample date) and the predicted value of the dependent variable $\hat{y}$ obtained from the multiple regression model is called the **error** or **residual**.

Error=Actual Value−Predicted Value

For the simple linear regression model, the standard error of the estimate measures the average vertical distance (the error) between the points on the scatter diagram and the regression line.

MACHINE LEARNING



The **standard error of the estimate**, is a measure of the standard deviation of the errors in a regression model. The standard error of the estimate is a measure of the average deviation of the errors, the difference between the $\hat{y}$-values predicted by the multiple regression model and the y values in the sample. The standard error of the estimate for the regression model is the standard deviation of the errors/residuals.

The standard error of the estimate tells us, on average, how much the dependent variable differs from the regression model based on the independent variables. The units of the standard error of the estimate are the same as the units of the dependent variable.