

# 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?

Ans- A) Least Square Error

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

Ans - A) Linear regression is sensitive to outliers

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

Ans - B) Negative

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

Ans- B) Correlation

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

C) Low bias and high variance

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

Ans - B) Predictive modal

7. Lasso and Ridge regression techniques belong to \_\_\_\_\_?

Ans- D) Regularization

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

Ans- D) SMOTE

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

Ans - A) TPR and FPR

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

Ans- B) False

11. Pick the feature extraction from below:

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?

Ans – A,B,C

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

**13.Explain the term regularization?**

- Regularization is one of the most important concepts of machine learning. It is a technique to prevent the model from overfitting by adding extra information to it.
- Sometimes ML model performs well with the training data but does not perform well with the test data. It means the model is not able to predict the output when deals with unseen data by introducing noise in the output, and hence the model is called overfitted. This problem can be dealt with the help of a regularization technique.
- This technique can be used in such a way that it will allow to maintain all variables or features in the model by reducing the magnitude of the variables. Hence, it maintains accuracy as well as a generalization of the model.
- It mainly regularizes or reduces the coefficient of features toward zero. In simple words, "In regularization technique, we reduce the magnitude of the features by keeping the same number of features."

## 14. Which particular algorithms are used for regularization?

There are mainly two types of regularization algorithms , which are given below:

- **Ridge Regression(L2 norm)**
- **Lasso Regression(L1 norm)**

### **Ridge Regression**

Ridge regression is one of the types of linear regression in which a small amount of bias is introduced so that we can get better long-term predictions.

Ridge regression is a regularization technique, which is used to reduce the complexity of the model. It is also called as **L2 regularization**.

In this technique, the cost function is altered by adding the penalty term to it. The amount of bias added to the model is called **Ridge Regression penalty**. We can calculate it by multiplying with the lambda to the squared weight of each individual feature.

A general linear or polynomial regression will fail if there is high collinearity between the independent variables, so to solve such problems, Ridge regression can be used.

It helps to solve the problems if we have more parameters than samples.

### **Lasso Regression:**

Lasso regression is another regularization technique to reduce the complexity of the model. It stands for **Least Absolute shrinkage and Selection Operator**.

It is similar to the Ridge Regression except that the penalty term contains only the absolute weights instead of a square of weights.

Since it takes absolute values, hence, it can shrink the slope to 0, whereas Ridge Regression can only shrink it near to 0.

Hence, the Lasso regression can help us to reduce the overfitting in the model as well as the feature selection.

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

Error present in regression equation does not mean its kind of a mistake or something like that. It rather takes into consideration the hidden features present which accounts for the deviation from the fit curve.