

# MACHINE LEARNING

1. Which of the following methods do we use to find the best fit line for data in Linear Regression?

**ANS:-** Least Square Error

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

**ANS:-** Linear regression is sensitive to outliers

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

**ANS:-** Negative

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

**ANS:-** Correlation

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

**ANS:-** Low bias and high variance

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

**ANS:-** Predictive modal

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

**ANS:-** Regularization

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

**ANS:-** SMOTE

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

**ANS:-** Sensitivity and Specificity

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

**ANS:-** False

**11.** Pick the feature extraction from below:

ANS:- Apply PCA to project high dimensional data

**12.** More than one options are correct, choose all the correct options:

ANS:-(a) We don't have to choose the learning rate.

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

**13.** Explain the term regularization?

ANS:- Regularization is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid over-fitting.

The commonly used regularization techniques are :-

1. L1 regularization
2. L2 regularization

- ❖ A regression model which uses L1 Regularization technique is called LASSO(Least Absolute Shrinkage and Selection Operator) regression.
- ❖ A regression model that uses L2 regularization technique is called Ridge regression.
- ❖ Lets have some examples of overfitting and underfitting models:-

Consider a simple example. you are trying to predict the score of students in the exam. We use a **number of books read** as a **feature** to predict.

This model will not learn anything new but it can find a few patterns but not enough to predict the score. This is called the **underfitting of the model** means you didn't provide enough data to understand the pattern. The model is **too simple and high bias**.

The below visuals should be easy to understand the green line is not a good fit.

Now consider You went to identify that adding more features could improve the prediction and you added 20 more **features like sleep time, no of the book he referred, no of labs he attended, no of hours of study**, etc. We see that the model has started to pick a pattern to get too complex because of more input features.

*Now our model has also learned data patterns along with the noise in the training data. This is called overfitting.*

When a model tries to fit the data pattern as well as noise then the model has a high variance that will be overfitting.

The above visuals show it **detected patterns well on training data** but will do poorly on the test set and not generalize.

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

ANS:- **There are 3 different types of Regularization algorithms. And they are:-**

- \* Ridge Regression.
- \* LASSO (Least Absolute Shrinkage and Selection Operator) Regression.
- \* Elastic-Net Regression.

Let's discuss these 3 algorithms in brief:-

### ***Ridge Regresssion***

Ridge Regression **solves the problem of overfitting** , as just regular squared error regression fails to recognize the less important features and uses all of them, leading to overfitting. Ridge regression adds a slight bias, to fit the model according to the true values of the data. Ridge regression is the method used for the analysis of multicollinearity in multiple regression data. It is most suitable **when a data set contains a higher number of predictor variables than the number of observations**. The second-best scenario is when multicollinearity is experienced in a set.

## ***Lasso Regression***

Lasso regression is also called Penalized regression method. This method is **usually used in machine learning for the selection of the subset of variables**. It provides greater prediction accuracy as compared to other regression models. Lasso Regularization helps to increase model interpretation.

## ***Elastic-Net Regression***

Elastic net is a hybrid of ridge regression and lasso regularization. Elastic net is always preferred over lasso & ridge regression because it solves the limitations of both methods, while also including each as special cases. So if the ridge or lasso solution is, indeed, the best, then any good model selection routine will identify that as part of the modeling process.

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

**ANS:-** The standard error of the regression is the average distance that the observed values fall from the regression line. Linear regression most often uses **mean-square error (MSE)** to calculate the error of the model. MSE is calculated by: 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.

### **Lets' discuss how is the error calculated in a linear regression model?**

- 1.measuring the distance of the observed y-values from the predicted y-values at each value of x;
- 2.squaring each of these distances;
- 3.calculating the mean of each of the squared distances.