# 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?
   1. Least Square Error B) Maximum Likelihood

C) Logarithmic Loss D) Both A and B

Ans. A) Least Square Error

1. Which of the following statement is true about outliers in linear regression?
   1. 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

1. A line falls from left to right if a slope is ?
   1. Positive B) Negative

C) Zero D) Undefined

Ans. B) Negative

1. Which of the following will have symmetric relation between dependent variable and independent variable?
   1. Regression B) Correlation

C) Both of them D) None of these

Ans. (B) Correlation

1. Which of the following is the reason for over fitting condition?
   1. 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

1. If output involves label then that model is called as:
   1. Descriptive model B) Predictive modal

C) Reinforcement learning D) All of the above

Ans. B) Predictive modal

1. Lasso and Ridge regression techniques belong to ?
   1. Cross validation B) Removing outliers

C) SMOTE D) Regularization

Ans. D) Regularization

1. To overcome with imbalance dataset which technique can be used?
   1. Cross validation B) Regularization

C) Kernel D) SMOTE

Ans. D) SMOTE

1. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses to make graph?
   1. TPR and FPR B) Sensitivity and precision

C) Sensitivity and Specificity D) Recall and precision

Ans. C) Sensitivity and Specificity

1. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.
   1. True B) False

Ans. B) False

1. Pick the feature extraction from below:
   1. Construction bag of words from a email
   2. Apply PCA to project high dimensional data
   3. Removing stop words
   4. Forward selection

Ans. A) Construction bag of words from a email

# In Q12, more than one options are correct, choose all the correct options:

1. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?
   1. We don’t have to choose the learning rate.
   2. It becomes slow when number of features is very large.
   3. We need to iterate.
   4. It does not make use of dependent variable.

Ans. A) & B)

13. In machine learning Overfitting is a model that models the training data too well. This happens when a model learns all the details and noise in the data upto a extent that it later negatively impacts the new data we will work upon. In other words we can say that our model is working well on the training data but is not able to work well on the unseen data. Therefore one of the most important task in machine learning is to avoid Overfitting.

One of the important techniques which limits Overfitting in machine learning is Regularization. It is a form of regression that regresses or shrinks the coefficient estimates towards zero. The technique of Regularization restricts or discourages the learning of a complex model so that we avoid the risk of Overfitting in the new data. It does so by adding a penalty or with Residual Sum of Squares (RSS) to the complex model.

Let us consider now a Simple linear regression equation:

Here Y represents the dependent feature Then,

Y is approximated to β0 + β1X1 + β2X2 + …+ βpXp

Here, X1, X2, …Xp are the independent features, and

β0, β1,…..βn represents the coefficients estimates for different variables, which describes the magnitude attached to the features.

In regression, our loss function is known as the residual sum of squares (RSS).

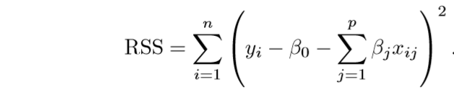


Fig. Cost Function For Simple Linear Regression

Now, this will adjust the coefficient estimates based on the training data. If there is noise present in the training data, then the estimated coefficients are not able to predict the future data.

Now here regularization will act and shrinks the learned estimates towards zero, by adding a loss function that can predict the accurate value of Y.

14.There are two types of regularization algorithms used:

* Ridge Regression
* Lasso Regression

### Ridge Regression- In this we introduce a small amount of bias, known as Ridge regression penalty so that we can get good predictions.

In this way the cost function of data is changed by adding the penalty term which has multiplied the lambda with the squared weight of each individual feature. Therefore, cost function now becomes:

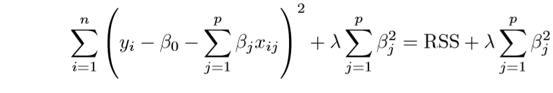


Fig. Cost Function for Ridge Regression

## Lasso Regression- In this the penalty term includes the absolute weights instead of a square of weights. Therefore, the cost function becomes:

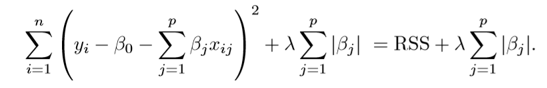


Fig. Cost Function for Lasso Regression

In this some of the coeﬃcient estimates are changed to zero which means there is a complete removal of some of the features.

15.The term error present in linear regression equation represents the total sum of the standard deviation observed in the regression line. This clearly explains or points out the difference of values observed between the observed and the expected(theoretical) data.

If there is an error observed in some set of data then it is said that the model is not completely accurate because there are variations in its observed and expected values. For example: let us consider linear regression equation such as:

*Y*=*αX*+*βρ*+*ϵ*

where: *α*,*β*=Constant parameters

*X*,*ρ*=Independent variables

*ϵ*=Error term​﻿

So, this is how error is represented in the equation. Now when the expected and observed value of Y values then the value of Error is not equals to zero, which clearly indicates that there are other factors which are influencing the value of Y.