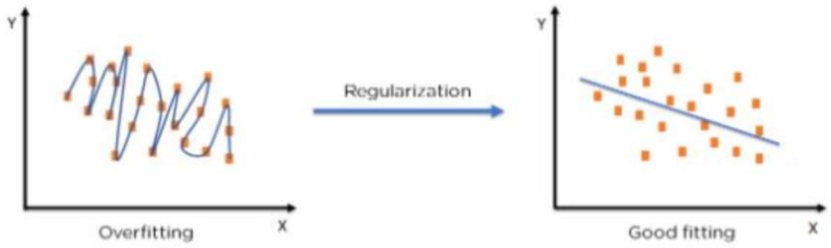
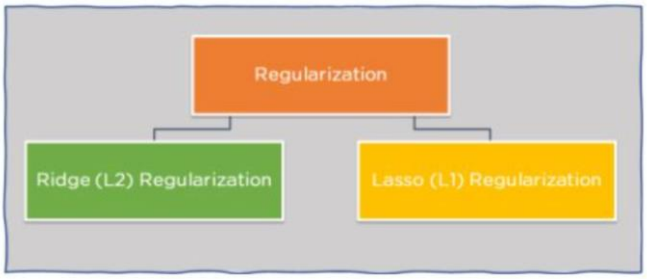
Q.1. Explain the term regularization?

Ans. Regularization is one of the most important concepts of machine learning. Regularizations are techniques used to reduce the error by fitting a function appropriately on the given training set and avoid overfitting. The word regularize means to make things regular or acceptable. Sometimes the machine learning model performs well with the training data but does not perform well with the test data.

Figure : Regularization on an over-fitted model



There are two main types of regularization techniques: Ridge Regularization and Lasso Regularization.



Q.2. Which particular algorithms are used for regularization?

Ans. Algorithms are used for regularization:

- Ridge Regression

  Ridge regression is a model tuning method that is used to analyse any data that suffers from multicollinearity. This method performs L2 regularization. When the issue of multicollinearity occurs, least-squares are unbiased, and variances are large, this results in predicted values being far away from the actual values.

- LASSO (Least Absolute Shrinkage and Selection Operator) Regression
  Lasso regression is a type of linear regression that uses shrinkage. Shrinkage is where data values are shrunk towards a central point, like the mean. The lasso procedure encourages simple, sparse models (i.e. models with fewer parameters)

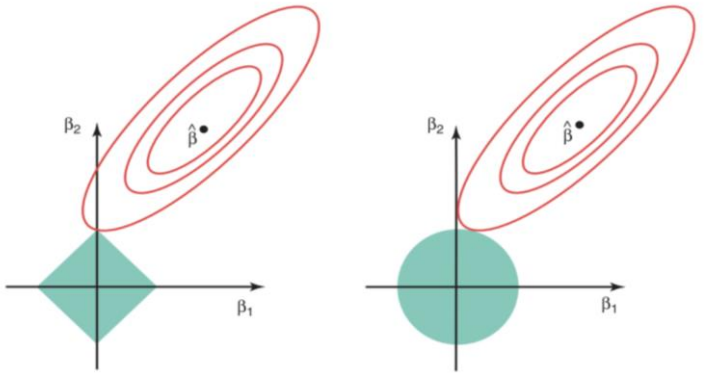


FIGURE 6.7. *Contours of the error and constraint functions for the lasso* (left) *and ridge regression* (right). *The solid blue areas are the constraint regions,* $|\beta_1| + |\beta_2| \leq s$ *and* $\beta_1^2 + \beta_2^2 \leq s$, *while the red ellipses are the contours of the RSS.*

Both Ridge and Lasso are regularisation techniques and helps in reducing the overfitting if a model. Both these methods shrink the coefficients and there is a difference in the way they do it.

Lasso is capable of making the coefficients to zero and Ridge can make them only very close to zero. So if variable selection is your objective, then you can go with Lasso. If you are trying to just reduce the overfitting, then you can go with either of these techniques

Q.3. Explain the term error present in linear regression equation?

Ans. In 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. Linear regression fits a line to the data by finding the regression coefficient that results in the smallest MSE.