**Problem 4: Using Gradient Descent for Ridge Regression Learning**

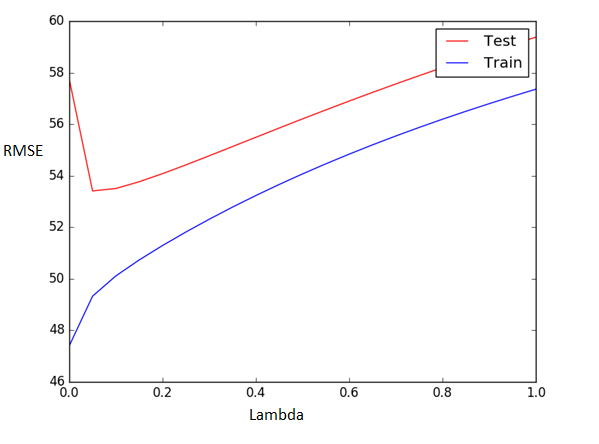
In *regressionObjVal* function, given data (X) and corresponding labels (y), and regularization parameter (λ), we calculated squared error value using the equation for regularized squared loss function as below:



Then, the gradient of squared error with respect to w is given by the formula:



The error value and its gradient are then used to minimize the weights for Ridge Regression.



*Figure 4*

Similar to problem 3, RMSE initially decreases steeply with increase in λ value. Due to underfitting, when the λ value increases, error also increases.

Using the gradient descent method provides satisfactory results when the number of features becomes large or in cases where finding the inverse of the covariance matrix is not possible.

On comparing the graph above for training and testing data for Ridge Regression using Gradient descent with that of simple Ridge Regression, we can infer that the error value (RMSE) doesn’t change much between both methods. The conclusion is that by observing both the graphs ,the optimal lambda values are approximately same for both the methods. i.e.

λ = 0.0599 for simple Ridge Regression and λ = 0.0499 for Ridge Regression using gradient descent.