CSE-574 Machine Learning

Programming Assignment 2

Classification and Regression

Spring 2017

University at Buffalo

Group Number: **22**

<u>Members:</u>

Hrishikesh Saraf	5020 5927
Nikita Konda	5020 7717
Shreya Nimje	5020 6520

Problem 1: Experiment with Gaussian Discriminators

Using the given training data set we trained the Gaussian Discriminators and evaluated the performance on the testing data set. The accuracy of the Linear Discriminant Analysis (LDA) on the testing data set was observed to be 97%. The accuracy of the Quadratic Discriminant Analysis (QDA) on the testing data set was observed to be 96%.

The following graph plots the discriminating boundaries for both Linear and Quadratic Discriminators.

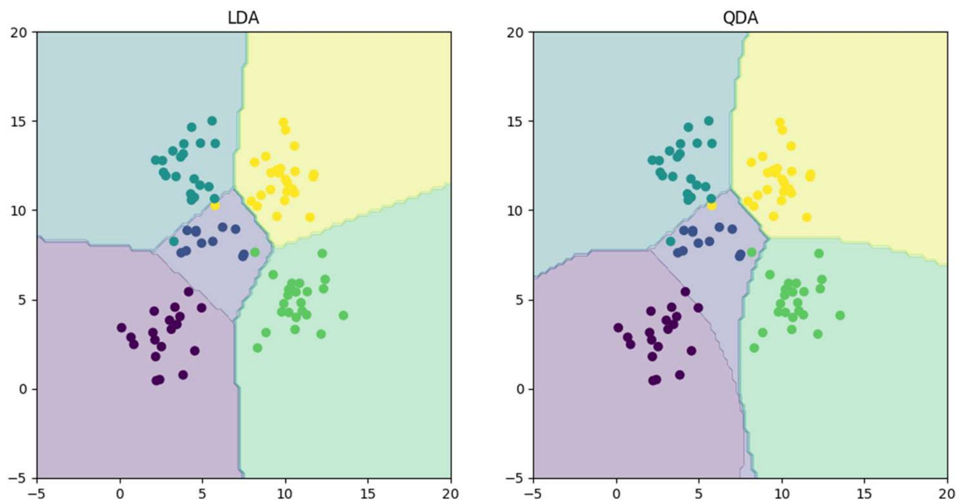

Fig 1: MSE values for Ridge Regression

We observed that the boundaries for both the graphs are different. This is due to the difference in the way we calculate covariance for the two discriminators. In LDA, the covariance is calculated for the entire training set. Whereas, in QDA for every class the covariance is calculated separately.

Problem 2: Experiment with Linear Regression

<u>MSE Values without using an Intercept:</u>

I. Training Data: 19099.446844570855
II. Testing Data: 106775.36155357839

<u>MSE Values Using an Intercept:</u>

I. Training Data: 2187.1602949303847
II. Testing Data: 3707.8401815379502

MSE value for Linear Regression using an intercept gives less error and hence is better as compared to the MSE without using an intercept.
Also MSE for training data is better than for testing data as training data gives low error.

Problem 3: Experiment with Ridge Regression

The MSE for training data was **2451.53** and testing data was **2851.33** using the Ridge Regression parameter lambda (**λ**) = 0.06.

The following graph shows the MSE for training dataset and test dataset plotted against varying values of **λ** from 0 to 1 in steps of 0.01.

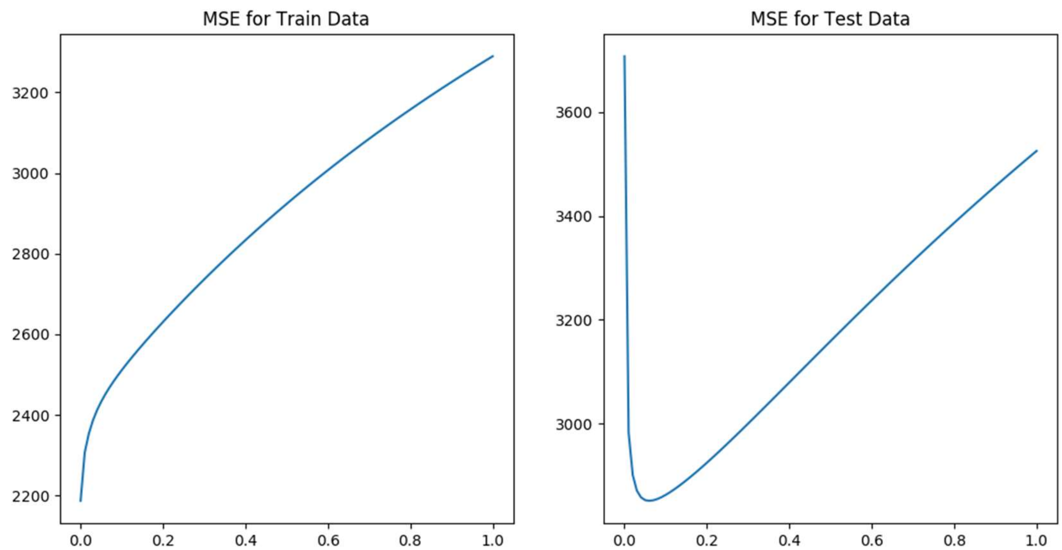

Fig 2: MSE values vs Regularization parameter (**λ**) for Ridge Regression

The following graph shows the comparison of relative magnitudes of weights learnt in Linear Regression using OLE and weights learnt in Ridge Regression. It was observed that in Linear Regression, magnitude of weights were in the range of -90000 to 80000 whereas in Ridge Regression it was in the range of -120 to 220. So, the magnitudes of weights learnt in Linear Regression were much higher than the magnitudes of weights learnt in Ridge Regression.

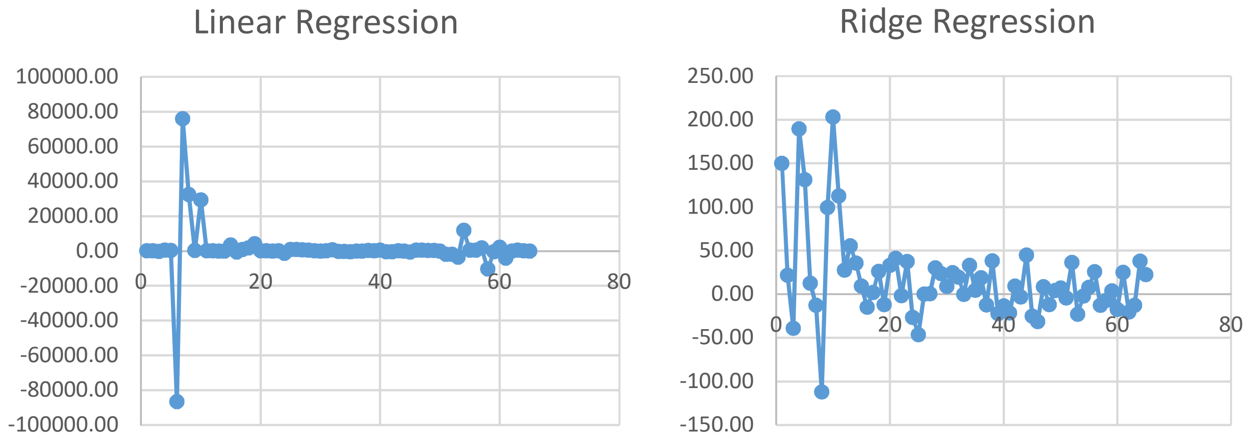

Fig 3: Magnitudes of weights learnt in Linear Regression vs Ridge Regression

Comparison of error on training and testing data in Linear Regression and Ridge Regression:

MSE	Training Data	Test Data
Linear Regression	2187.1602949303847	3707.8401815379502
Ridge Regression	2451.52849064	2851.33021344

It was observed that the MSE for linear regression was more than MSE for ridge regression of testing data set whereas for training data set it was vice versa.

The optimal value of lambda (**λ**) is **0.06** because the MSE for testing data was observed to be minimum for this particular value of lambda.

Problem 4: Using Gradient Descent for Ridge Regression Learning

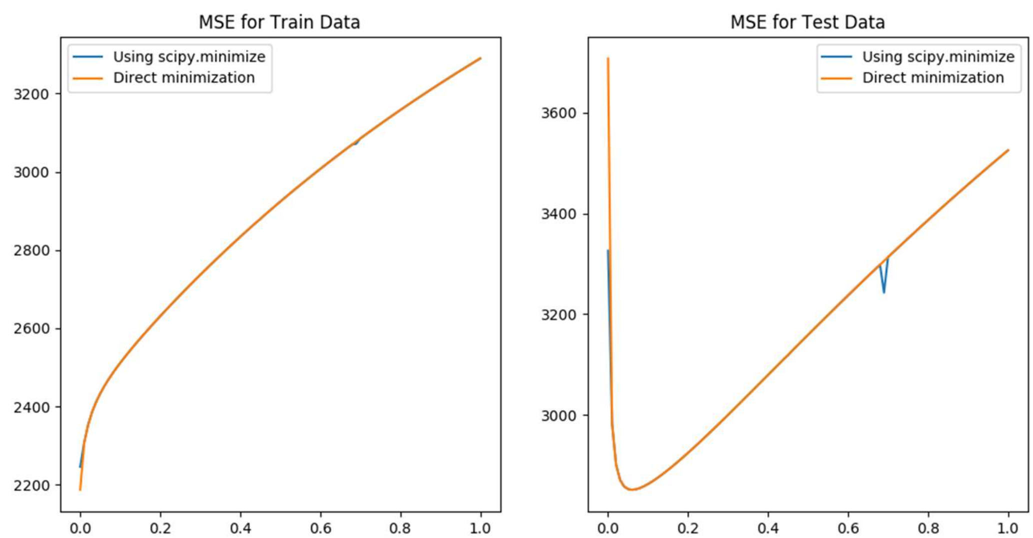

Fig 4: MSE values for Ridge Regression Using Gradient Descent

Comparing the graph of the MSE obtained for Ridge Regression i.e. direct minimization and MSE for Ridge Regression using Gradient Descent, we observe that both the curves are almost similar or very minute changes are seen with number of iterations as 100. As the Lamba value increases, we get a smooth curve for both Ridge Regression with and without using Gradient Descent but the curve is slightly better for Ridge Regression without using gradient descent i.e. direct minimization.

Problem 5: Non-linear Regression

The following graph plot of MSE on training and testing data was obtained by training ridge regression weights using the non-linear mapping of data by varying **p** from 0 to 6 with and without regularization.

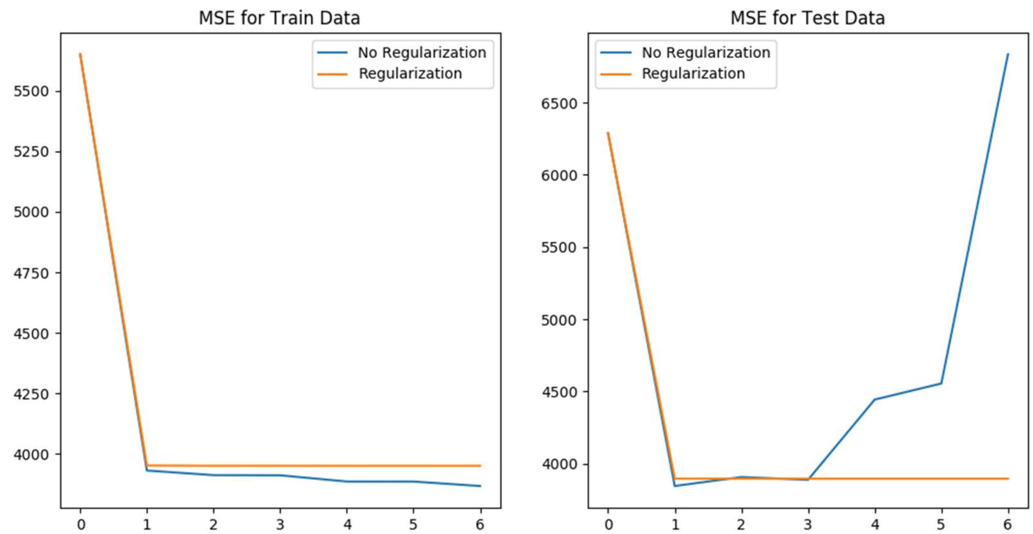

Fig 5: MSE values for Ridge Regression Using Gradient Descent

On comparing the graph for both the values of **λ=0** (No regularization) and **λ-opt=0.06** (Regularization), we observed that for the training dataset the MSE for λ=0 decreased with an increase in p whereas for the testing dataset the MSE for λ=0 increased with an increase in p. But with regularization λ-opt=0.06 the MSE was constant after a particular point for both training and testing set.

The optimal value of p in terms of testing data with **λ=0** or **λ-opt=0.06** is 1 because the MSE was observed to be minimum when p was set to 1 in both the setting.

Problem 6: Interpreting Results

We implemented in all 4 methods for predicting the diabetes level using regression.

<u>Linear Regression</u>: The MSE for training data was less compared to other methods but the MSE for testing data was high.

<u>Ridge Regression:</u> The MSE for testing data was the least as compared to all the other methods at λ-opt=0.06.

<u>Ridge Regression using Gradient Descent:</u> The method of minimizing the error function was the only difference between the previous Ridge regression and this method. So, the results were almost the same as previous with 100 number of iterations.

Non-Linear Regression: Amongst all the values of p, the MSE was lowest when the value of p was set to 1, but still the MSE was the highest as compared to other 3 methods.

Summarizing all the methods for regression as shown in the following table, we see that the MSE for testing set is minimum while using **Ridge Regression at λ-opt=0.06**

MSE	Training Data	Test Data
Linear Regression	2187.1602949303847	3707.8401815379502
Ridge Regression at λ=0.06	2451.52849064	2851.33021344
Ridge Regression using Gradient Descent at λ=0.06	2451.53096557	2851.33288279
Non-Linear Regression at p=1 and λ=0	3930.91540732	3845.03473017
Non-Linear Regression at p=1 and λ=0.06	3951.83912356	3895.85646447

So, for anyone using regression to predict the diabetes level, **Ridge Regression** at **λ-opt=0.06** will be best solution since it has the lowest testing error as compared to all the other methods that was implemented in this assignment.