CSE 574:Introduction to Machine Learning
Programming Assignment 2:Classification and Regression

Group Members (Group 26):
1. Nischala Manjunath
2. Reshma Raghavan
3. Sai Srinath Sundar

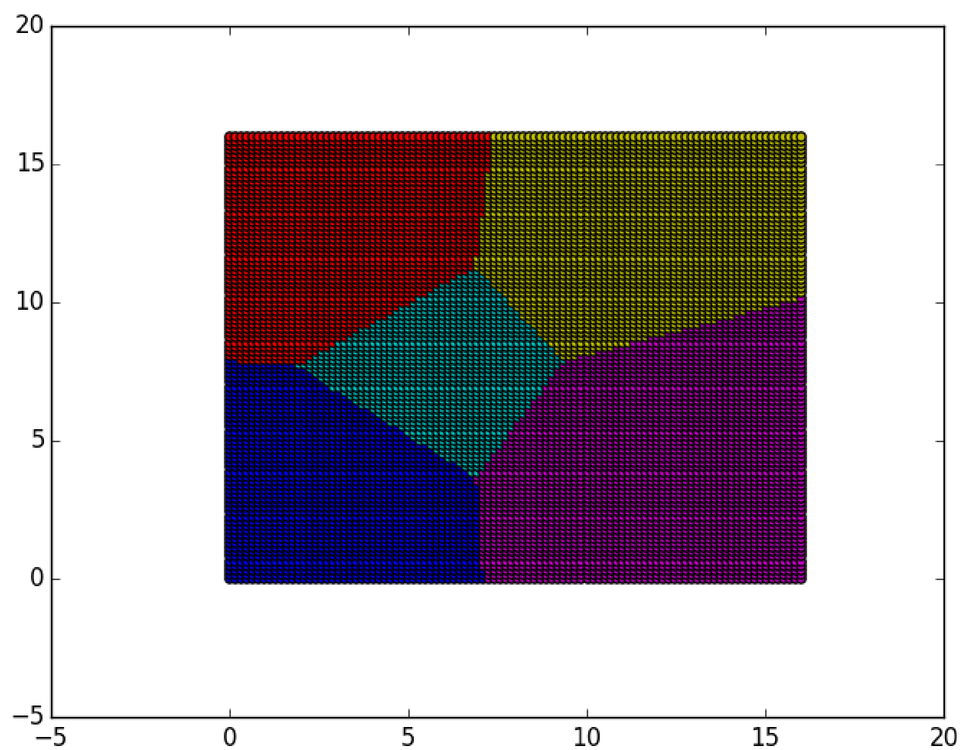
Overview

We had performed various classification and regression techniques using the datasets "sample.pickle" and "diabetes.pickle" containing both training and test data. The report contains detailed descriptions and results with respect to each technique.

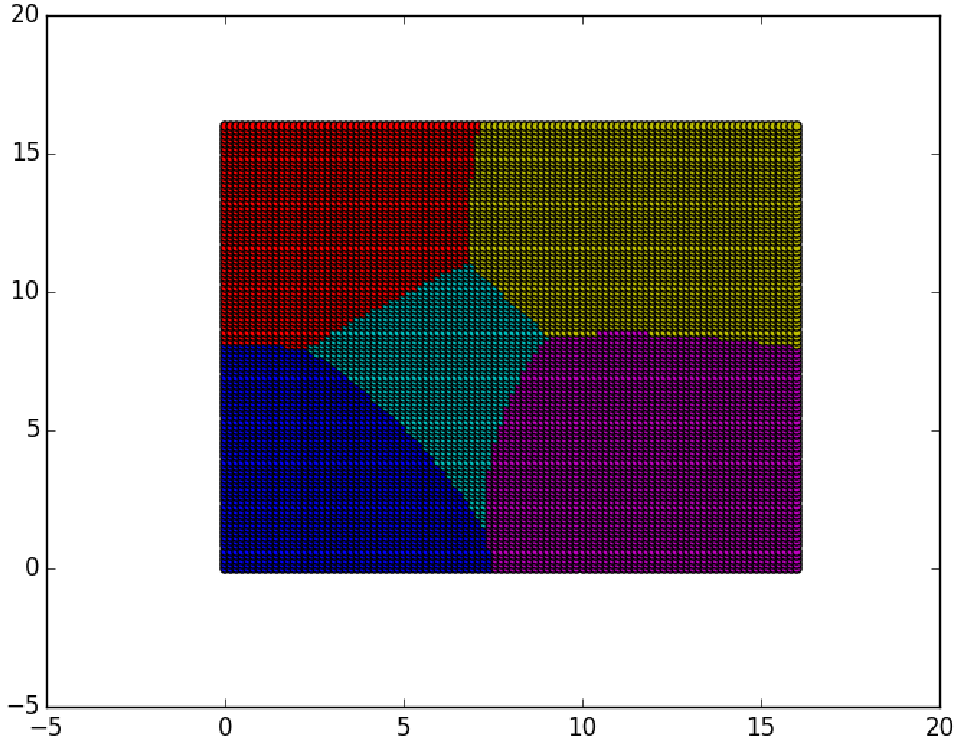
Problem 1

The sample dataset was trained with both LDA and QDA and the accuracy obtained for both methods are given as:
LDA Accuracy = 97.0
QDA Accuracy = 97.0

The Discriminating Boundary for Linear Discriminant Analysis is given as:



The Discriminating Boundary for Quadratic Discriminant Analysis is given as:



The difference observed in the discriminating boundaries between LDA and QDA is the boundaries for LDA are straight lines as opposed to curved boundaries in the case of QDA. From this we infer that QDA offers more flexibility in fitting data within boundaries rather than conforming to the rigid straight lines that are a characteristic of LDA.

Problem 2

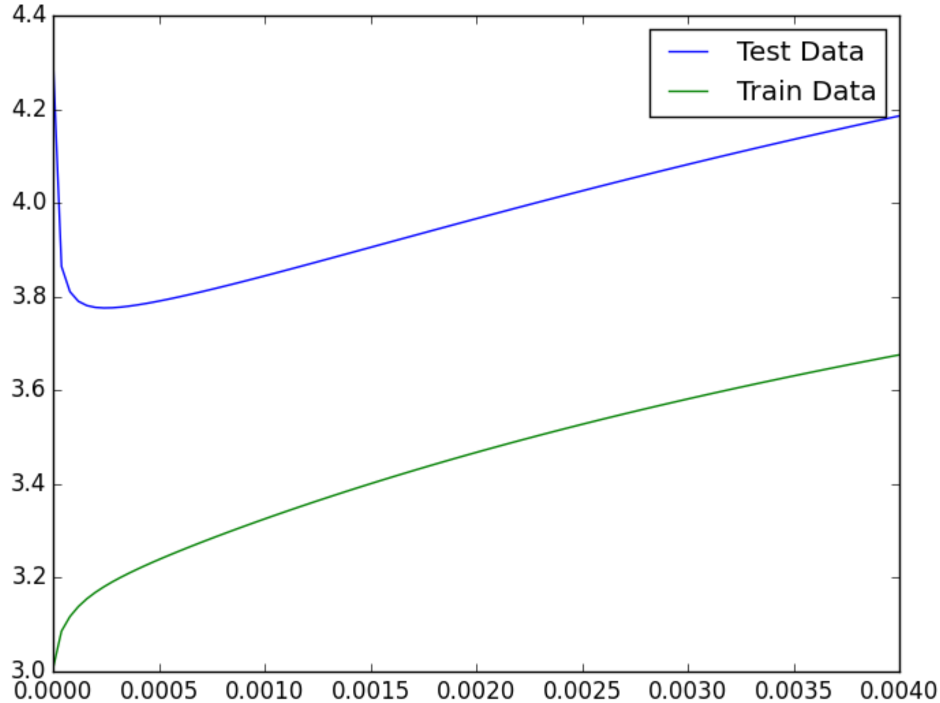
The RMSE for the training and test data was calculated for two cases: without using an intercept and with using an intercept. The observations are given below as:

RMSE without intercept: 23.1057743385
RMSE with intercept: 4.30571723473

From the given observations we can infer that the RMSE with the intercept in this case is far lesser than the RMSE without intercept, which implies that finding the RMSE with the intercept is a more accurate method in this case as the root mean square error is reduced by a ratio of about 5.

Problem 3

The plot of the errors on different values of Lambda for train and test data is given below:



From the given plot we can see that the optimal value of λ is approximately 0.0002. This implies that only a certain amount of regularization is necessary for optimal performance, and as λ increases, it incrementally constrains the parameters causing increased error.

On comparing the weights learned in Ridge Regression to OLE Regression, the relative magnitudes of the weights used in Ridge Regression is far greater than that used in OLE Regression. A sample of the first 4 rows of weights of both are given below to illustrate the difference in magnitude.

OLE Regression sample weights(w[0]-w[4]):

```
[ -4.12173302e+02][ -3.45940349e+02][ 5.78814085e+02][  5.89243804e+01]
```

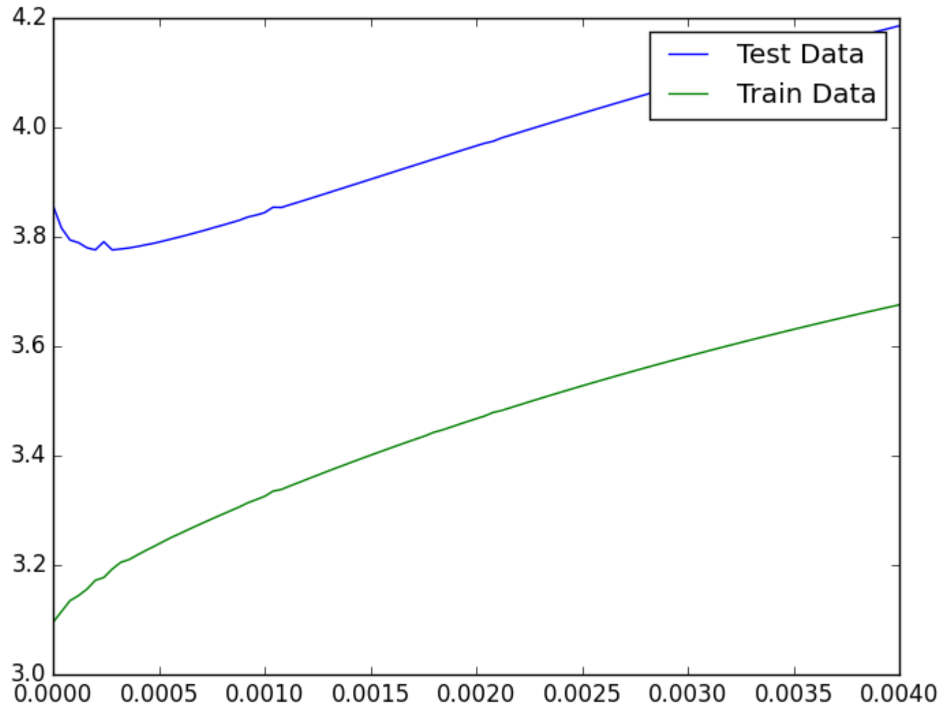
Ridge Regression sample weights(w[0]-w[4]):

```
[ -42.11979795]  [ -43.62736092]  [ 173.67280031]  [ 134.4406914 ]
```

The errors on train and test data are nearly identical with no regularization (λ=0) involved. With the optimal regularization, the error for Ridge Regression is lesser than that for OLE Regression. However as the regularization parameter increases,so does the error.

Problem 4

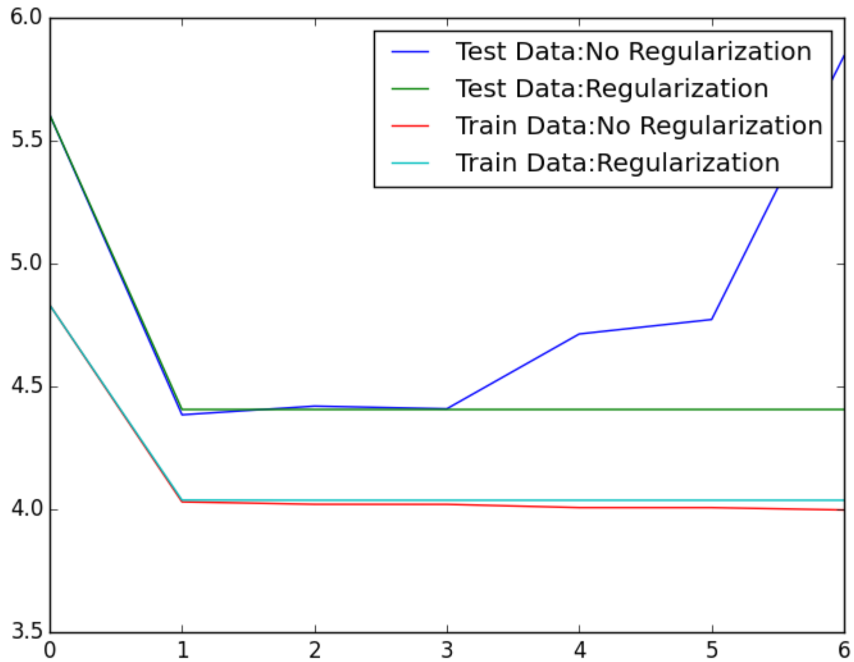
The plot of the errors on train and test data using gradient descent by varying Lambda is given below:



A comparison between the results of problem 4 and problem 3 was carried out and we can infer that the slope of the error for the train data using Gradient Descent is much lesser than the slope of the error on problem 3 with a non zero intercept in the case of Gradient Descent.

Problem 5

The optimal value of p for both values of Lambda were plotted and compared, given below:



The optimal values of p for the test data is (1,2,3) for both with and without regularization. With regularization, higher values of p give the same amount of test error.

The error on the train data is naturally lower than that on the test data. It must also be noted that regularization does not play a major part in the training data. In fact, going by the results obtained, regularization results in a marginally higher error for the training data.

Problem 6

Generically speaking, one cannot say with definite assertion that one technique is better than the other. It often is a factor of the type of the dataset and size of the dataset.

From the given techniques and datasets, if we were to recommend using regression for predicting diabetes levels using the input features, Ridge Regression would have an edge over Linear Regression provided the regularization parameter is around its optimal value. In this assignment, Ridge Regression seems to have an edge over non-linear regression as well.

While Linear Discriminant Analysis and Quadratic Discriminant Analysis give the same amount of accuracy in Classification, QDA seemed to fit the data better within its boundaries.

It must be noted that all these recommendations are made ignoring the computational complexity of the various methods. QDA is computationally more complex since it involves maintaining a covariance matrix for each class. This can cause a lot of overhead if the number of classes is large. Ridge Regression is also computationally expensive as it involves repeated Inversion and Transposition of matrices. Gradient Descent considerably reduces the overhead involved with Ridge Regression.