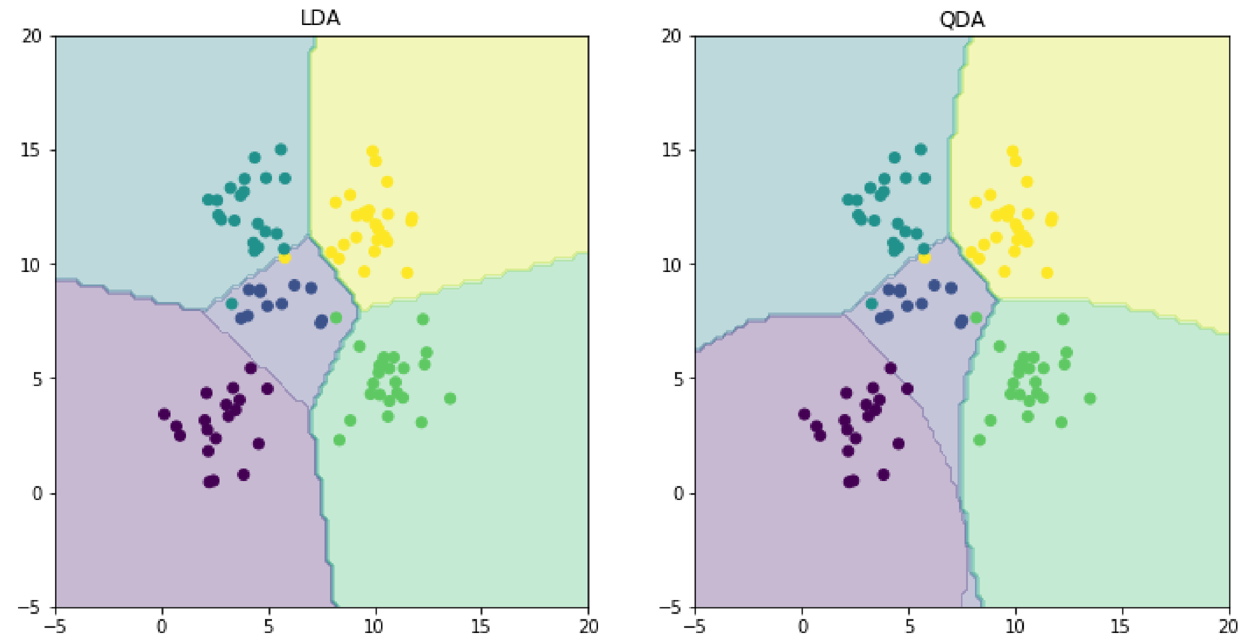
CSE-574
PROGRAMMING ASSIGNMENT-1
Project Report

Submitted by: Shreyas Kulkarni(sk385), Soham Gupte(sohamgup)

Report 1 :
Accuracies:

```
LDA Accuracy = 0.97
QDA Accuracy = 0.96
```



<u>Decision Boundaries in LDA and QDA</u>

The main difference between the LDA and the QDA decision boundaries is that the LDA decision boundary is linear because it assumes equal variance in observations in all classes. The QDA decision boundaries are quadratic as they are defined by quadratic equations. They are not as straight as the LDA decision boundaries as we compute different covariance matrices for each one of the classes and because the variability in observations in each of the classes is not the same.

Report 2:
Mean Squared Errors:

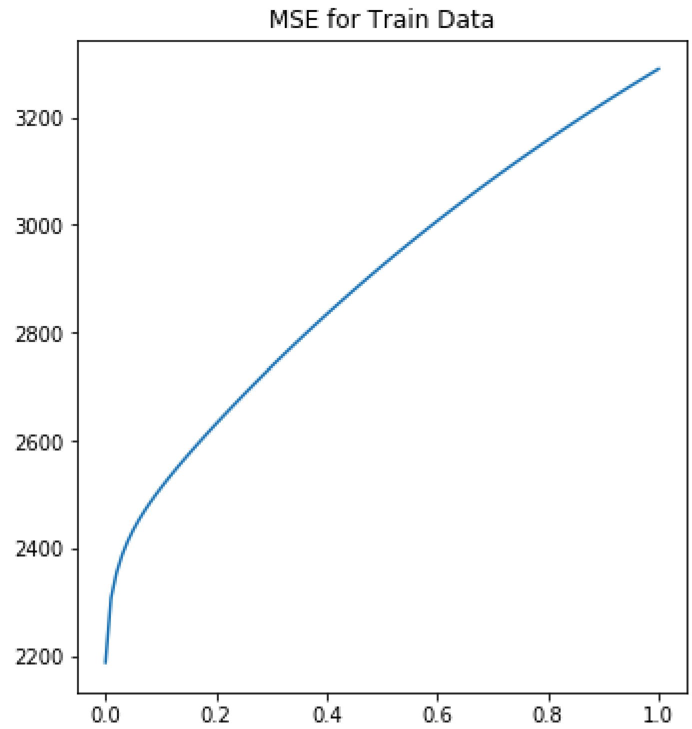
MSE without intercept =106775.361555
MSE with intercept = 3707.84018132

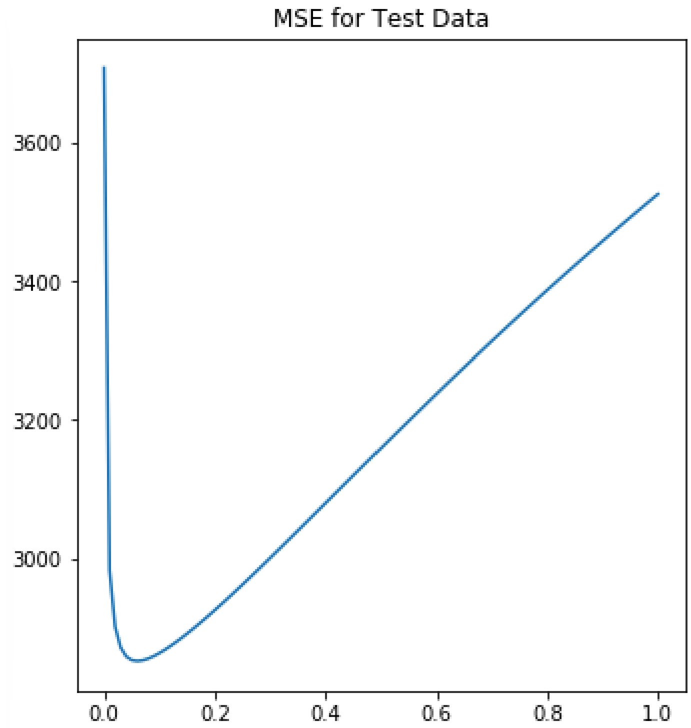
Clearly, the MSE dramatically reduces(almost by a factor of 30)when we add an intercept which is analogous to adding a bias term i.e. add a dimension to the data. Using an intercept based linear regression is better and helpful when predicting classes.

Report 3:

To avoid overfitting, we use L2 regularisation with linear regression which is called as ridge regression. We add square of magnitudes of coefficients multiplied by lambda in the result. The extent of shrinkage in weights of coefficients depends on the lambda values.
To compare w values learned from linear and ridge regression, we can observe that, output is unbiased in both cases. But, variation is more in linear regression. Bias- Variation trade-off is more balanced in ridge regression.

Below is the train error for different values of lambdas. Train error increases with the lambda values for training data.

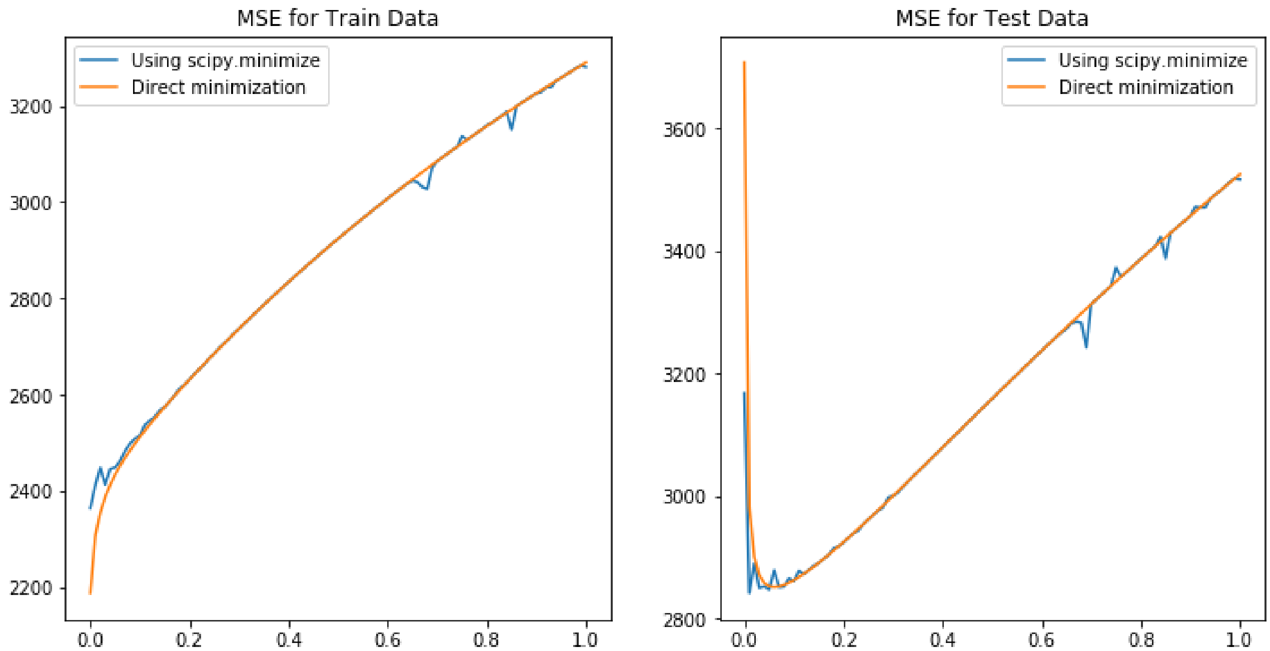




But this is not the case for test data. Error goes on decreasing with lambda values,becomes minimum at optimal lambda value and again takes sharp increasing curve after that value.

Optimal value of lambda is 0.06, because test error is minimum at that value with minimum variation of coefficients.

Report 4:

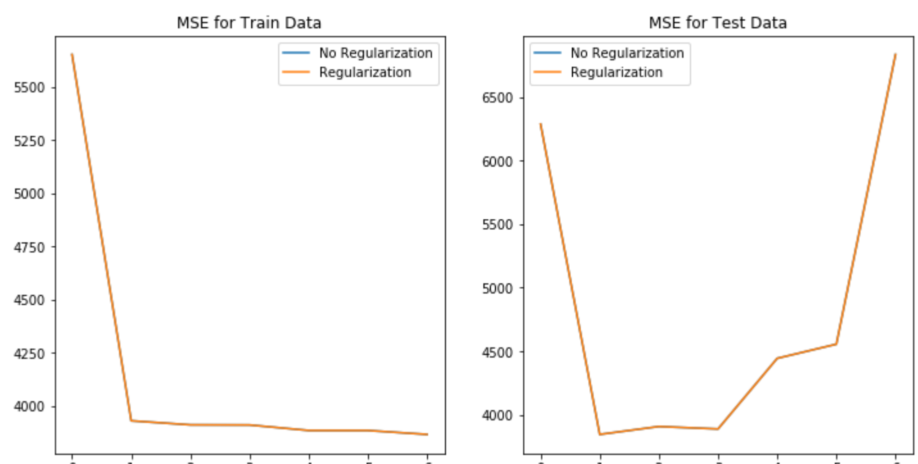


While not too dissimilar, it is clear the curve obtained in gradient descent based learning is not as smooth as the curve obtained in problem 3. The presence of spikes in the curve indicates that there are a few outliers and that the weights do not converge as quickly as normal ridge regression based learning. When the size of the training set is small both in terms of number and features, the dimensions of the matrices involved in the estimation of W will be small, thus reducing the computation time and overhead required for learning the weights. In such a case normal ridge regression based learning may still be used instead of gradient descent based learning as we can obtain a much smoother curve with approximately the same time overhead.

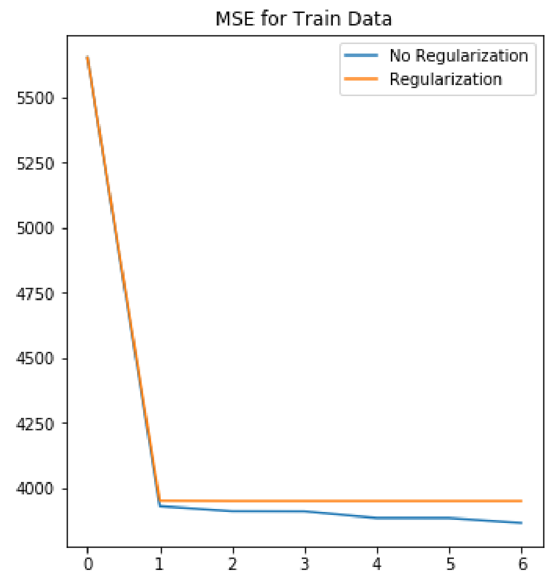
Report 5:

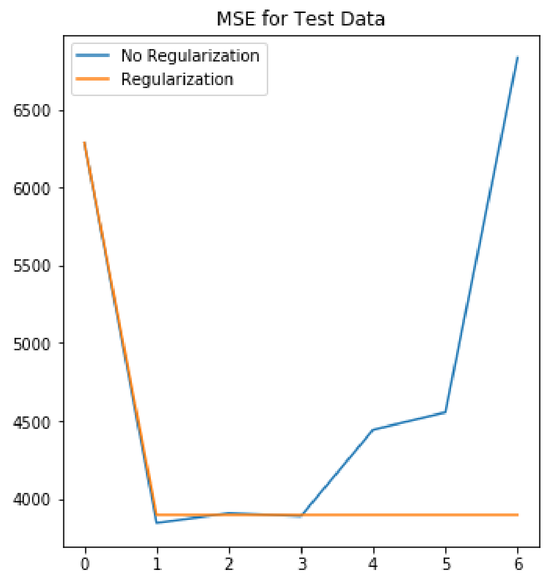
Variation in Mean Squared Error with p from 0 to 6

lambda=0



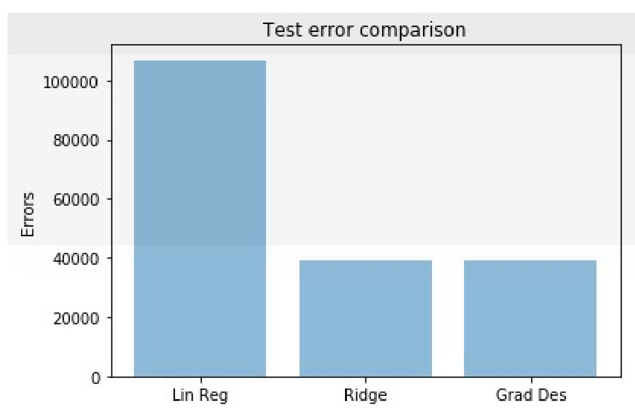
lambdopt=0.06





For training data:
As the order of the polynomial increases the squared error for the ridge regression learning(i.e. training error) decreases. The same holds true for both lambda=0 as well as lambdopt. So p=6 is the optimal value of p in this case.
For test data:
With no regularization, the optimal p value is p=1, in both cases. However, the curve is not as smooth as when regularization is used. In the curve for non-regularization the error suddenly increases after p=3. This is due to overfitting of the model when higher degree of polynomial is used.

Report 6:



From the graphs shown above, we can infer that:
Even though train errors of the methods are in the increasing order, test error decreases from Linear regression, ridge regression with optimal value to gradient descent. As we know, train error is not a true estimation of the method, test error actually reflects efficiency and accuracy of the method.

For polynomial of higher order, non-linear regression causes problem of over-fitting of the test data. So, consequently test error increases drastically after p=3.

So we can conclude that, gradient descent is the best method to predict diabetes level.