# CSE 574 INTRODUCTION TO MACHINE LEARNING

APRIL 11, 2017

## PA2: CLASSIFICATION AND REGRESSION

SUBMITTED BY: ABHISHEK YEOLA

ROSHAN BABY REJI

# CSE574 Machine Learning Programming Assignment 2

Group 6
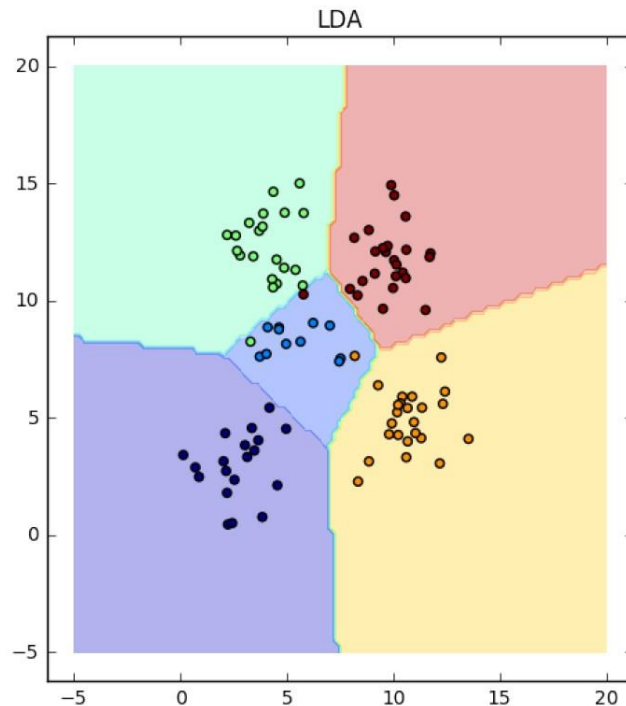Abhishek Yeola
Roshan Baby Reji

April 10, 2017

## Contents

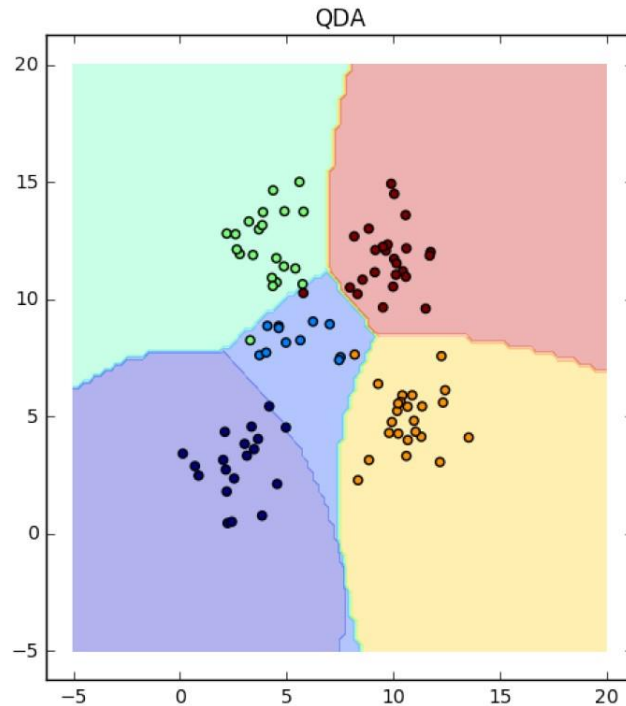## 1    Problem 1: Experiment with Gaussian Discriminators

Accuracy of LDA = 97%
Accuracy of QDA = 96%

**Discriminant boundary for LDA**



**Discriminant boundary for QDA**

QDA

The linear discriminant analysis (LDA) can only learn linear boundaries while the quadratic discriminant analysis (QDA) can learn quadratic boundaries which are more flexible. The LDA classifier assumes that each class has a normal distribution with a separate mean vector for the classes and a common covariance matrix. QDA estimates a separate covariance matrix for each class.

## 2  Problem 2: Experiment with Linear Regression

**Case 1:** Without using an intercept
MSE for training data: 19099.45
MSE for test data: 106775.36

**Case 2:** With using an intercept
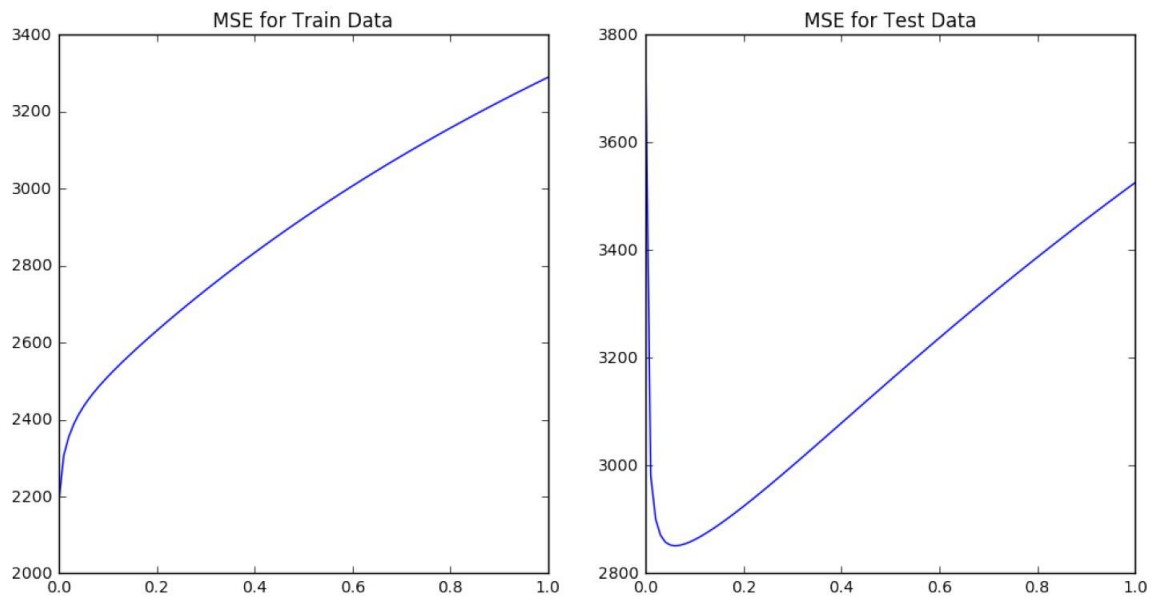MSE for training data: 2187.16
MSE for test data: 3707.84

Using an intercept for linear regression causes a significant decrease in the overall MSE. Thus linear regression using an intercept is a more accurate method.
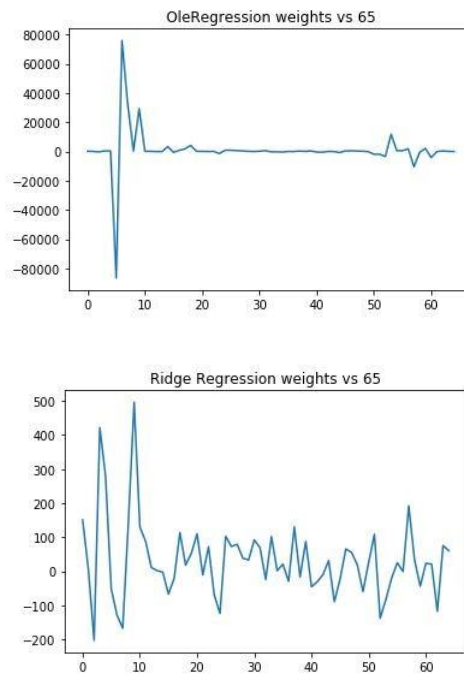
## 3  Problem 3: Experiment with Ridge Regression

Ridge regression using *testOLERegression* function. Data with intercept.

**Plot of $\lambda$ vs. error**

**Comparison of weights for OLE and Ridge Regression**





From this we can see that the weights for OLE regression are much higher that those of Ridge Regression. This is because the regularization factor in Ridge Regression favors low weights.

Minimum MSE for test data: 2851.33
Corresponding MSE for train data: 2451.53
Corresponding optimal $\lambda$: 0.06

# 4    Problem 4: Using Gradient Descent for Ridge Regression

$\lambda$ range: 0 to 1 $\lambda$ interval: 0.01 maximum number of iterations: 40 optimum $\lambda$ from problem 3: 0.06 MSE for train data at optimum $\lambda$: 2456.44

MSE for test data at optimum $\lambda$: 2852.47

**Plot $\lambda$ vs. error**
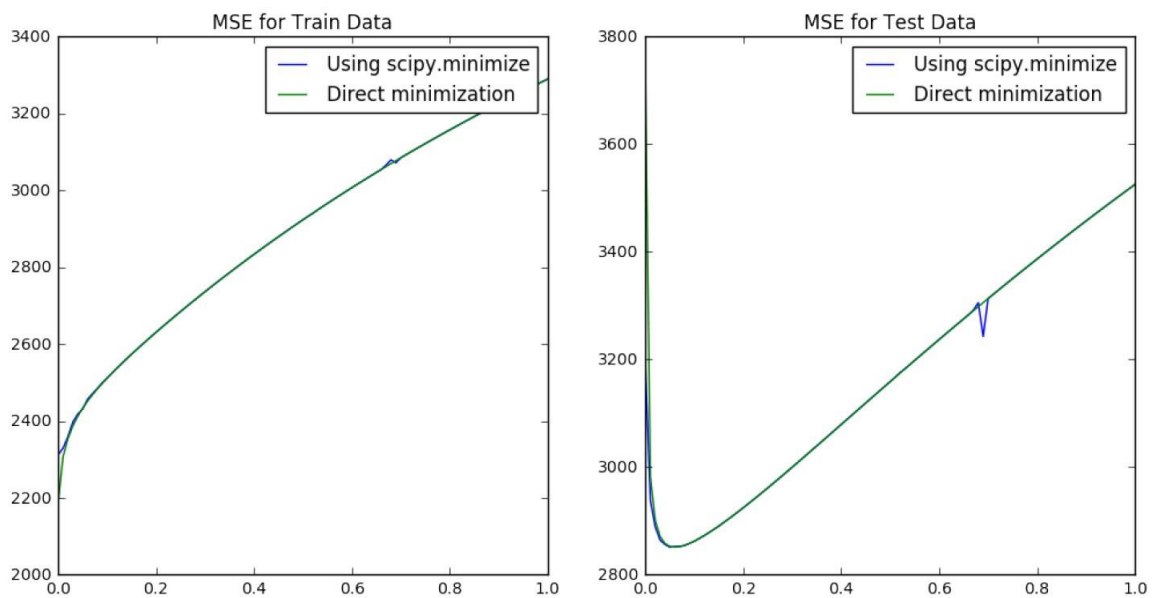


From the figure we can see that the results obtained in problem 3 are very similar to those obtained from problem 4. Since gradient descent is an iterative method, increasing the number of iterations will lead to better convergence. Using gradient descent to minimize the loss function does not require the calculation of matrix inverse. Thus it improves the accuracy of solution for higher dimensionality of data.

# 5    Problem 5: Non-linear Regression
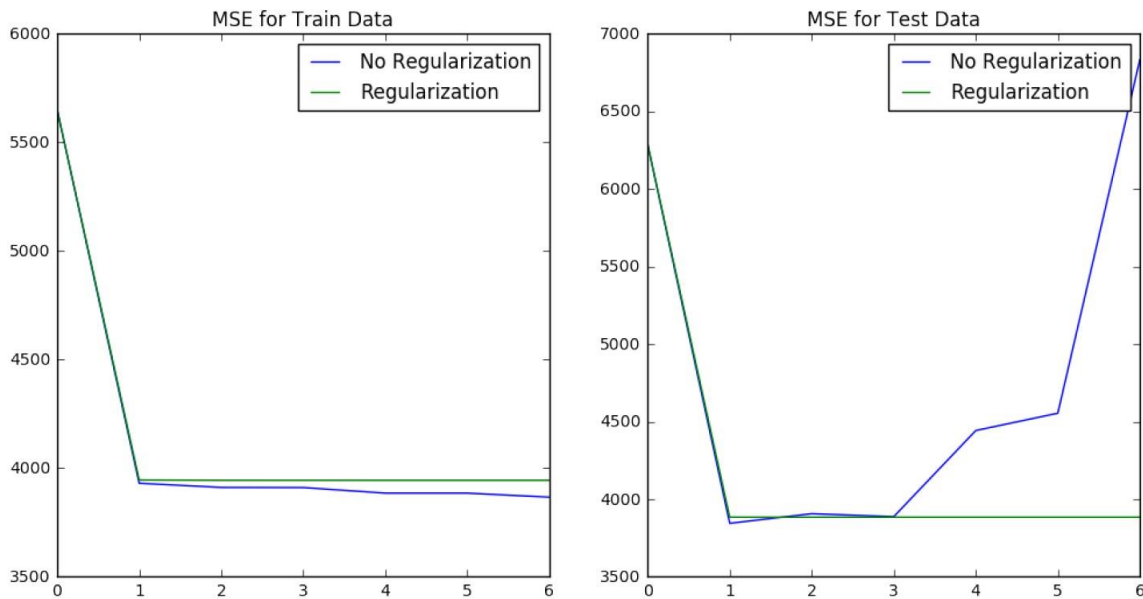
**Plot p vs. MSE**

Figure shows the MSE on the training set for different higher order polynomials of the input features. The degree of the polynomials varies from 0 to 6. The prediction error has been computed for two cases: without regularization, and with regularization with optimal value of $\lambda = 0.06$ from problem 3.

For training data, the MSE decreases when increasing the polynomial degree. This is due to the fact that we train using the same data set, therefore higher order curves will fit the data points more accurately and reduce the error.

For test data, when we do not use regularization, the error is minimum at p=1. Then it slowly increases until p=3, after which it rapidly increases. This is because with higher order polynomials we have overfitting which causes the training error to decrease. When we use regularization, the error is almost same for p=1 to p=6. This is due to the fact that the regularization term penalizes high weights. Therefore, even with high order polynomials, the correspondent weights with be very low.

For test data:

Optimum p for ($\lambda = 0$) = 1;

Optimum p for ($\lambda = 0.06$) = $1 \leq p \leq 6$;

# 6  Problem 6

| Problem | MSE training | MSE testing |
|---|---|---|
| OLE with intercept | 2187.16 | 3707.84 |
| OLE without intercept | 19099.45 | 106775.36 |
| Ridge regression (Lambda = 0.06) | 2451.53 | 2851.33 |
| Ridge regression – Gradient Descent (Lambda = 0.06) | 2456.44 | 2852.47 |
| Non Linear without regularization (Lambda = 0.06) | 3930.96 | 3845.03 |
| Non liner with regularization  (Lambda = 0.06) | 3950.68 | 3895.58 |

The MSE calculated for different methods is shown in the table above. The test MSE is used as a standard for measuring the performance of various regression methods. From the table, we can see that without using an intercept the error is very high. The MSE given by OLE is higher than that given by Ridge Regression. Within ridge regression the gradient descent method is a better option for higher dimensionality data set as it avoids the calculation of matrix inverse. Non Linear regression provides a good fit on the training data but may lead to overfitting problems. Also the error on test data is almost the same for p=1 to p=6.

So based on this data we recommend using Ridge regression without gradient descent for small data set and Ridge regression with gradient descent for large data set for the problem of predicting diabetes level using the given input features.