GROUP 18 Report

Department of Computer Science and Engineering

University at Buffalo, Buffalo, NY 14260

Hemanth Kumar Mutta

[hmutta@buffalo.edu](mailto:hmutta@buffalo.edu)

Desireddy Sai Sankeerthana

[saisanke@buffali.edu](mailto:saisanke@buffali.edu)

Sai Sandeep Lankisetty

[slankise@buffalo.edu](mailto:slankise@buffalo.edu)

**REPORT 1: Experiment with Gaussian Discriminators**

1. Linear Discriminant Analysis(LDA):

If we choose the class conditional densities in a special way, we will see that the resulting posterior over classes is a linear function of x, i.e., log p(y = c|x; θ) = wTx + const, where w is derived from θ. Thus the overall method is called linear discriminant analysis or LDA. 1

**p(x|y = c; θ)p(y = c; θ)**

**p(y = c|x; θ) = ----------------------------------------**

**∑ c ,p(x|y = c’; θ) p(y = c 0 ; θ)**

The term p(y = c; θ) is the prior over class labels, and the term p(x|y = c; θ) is called the class conditional density for class c.

Quadratic Discriminant Analysis (QDA):

**log p(y = c|x, θ) = log πc − 1 2 log |2πΣc| − 1 2 (x − µc ) TΣ −1 c (x − µc ) + const**

This is called the discriminant function. We see that the decision boundary between any two classes, say c and c 0 , will be a quadratic function of x. Hence this is known as QDA.

Accuracy on test data details as follows:

Accuracy of the test data for the implementation of two functions that are LDA Test and QDA Test where LDA Test is for Linear Discriminant Analysis and QDA Test is for Quadratic Discriminant Analysis

|  |  |  |
| --- | --- | --- |
|  | LDA | QDA |
| Accuracy | 97.00 | 96.00 |

Plot the discriminating boundary for linear and quadratic discriminators

**LDA QDA**

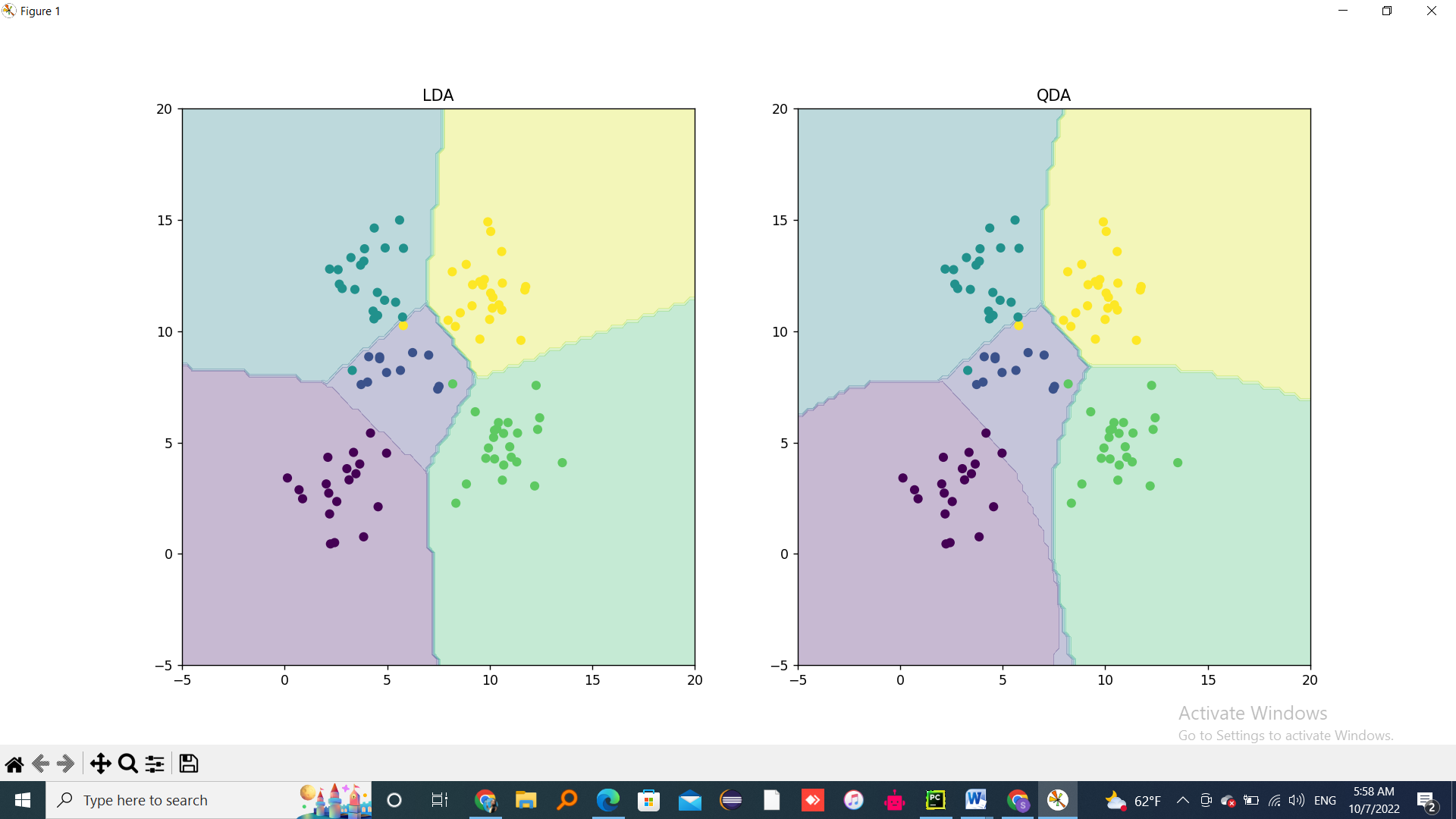
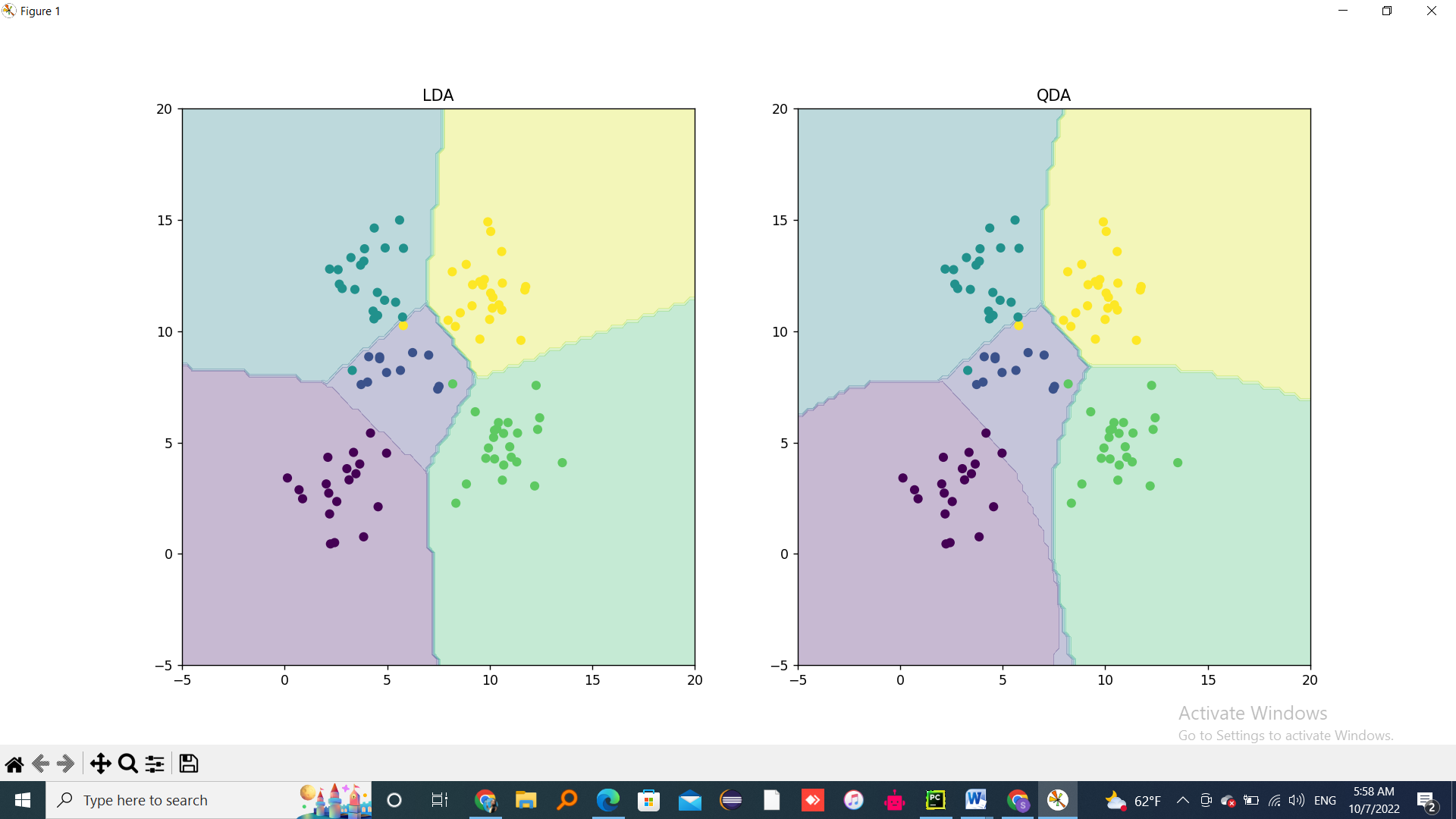
 

Figure 1:LDA and QDA

* LDA has a common covariance matrix. So, a covariance matrix that is common to all classes in a data set.
* As LDA has covariance matrix same for all classes that forces all the boundaries to be in straight lines
* QDA on other side, Observation of each class is drawn from a normal distribution.
* QDA assumes that each class has its own covariance matrix .hence , they don’t cancel lead to non-linear boundaries

**REPORT 2: Experiment with Linear Regression**

linear regression, which is a very widely used method for predicting a real-valued output (also called the dependent variable or target) y ∈ R, given a vector of real-valued inputs (also called independent variables, explanatory variables, or covariates) x ∈ R D. The key property of the model is that the expected value of the output is assumed to be a linear function of the input,

E [y|x] = wTx,

which makes the model easy to interpret, and easy to fit to data.

Table values for Mean Squared Error

|  |  |  |
| --- | --- | --- |
| MSE Values | Test Value | Train Value |
| With intercept | 3707.8401817 | 2187.16029493 |
| Without intercept | 106775.36150046 | 19099.44684457 |

The Linear Regression model with intercept is the better one .Intercept allows the model to passes more accurately through the data and hence provides better results.

**REPORT 3: Experiment with Ridge Regression**

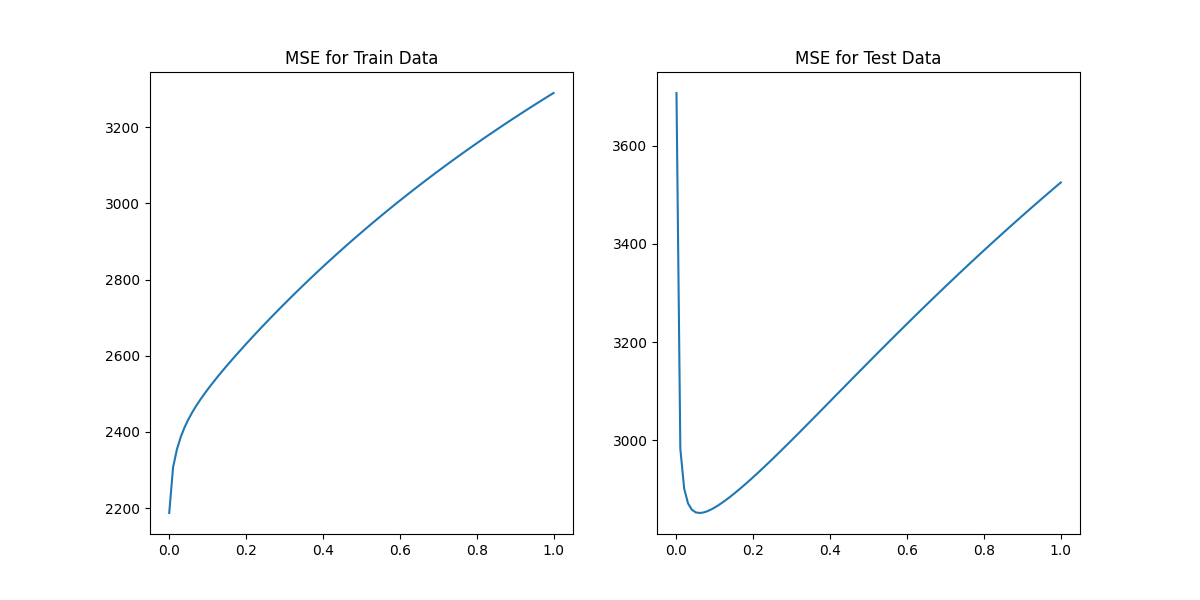
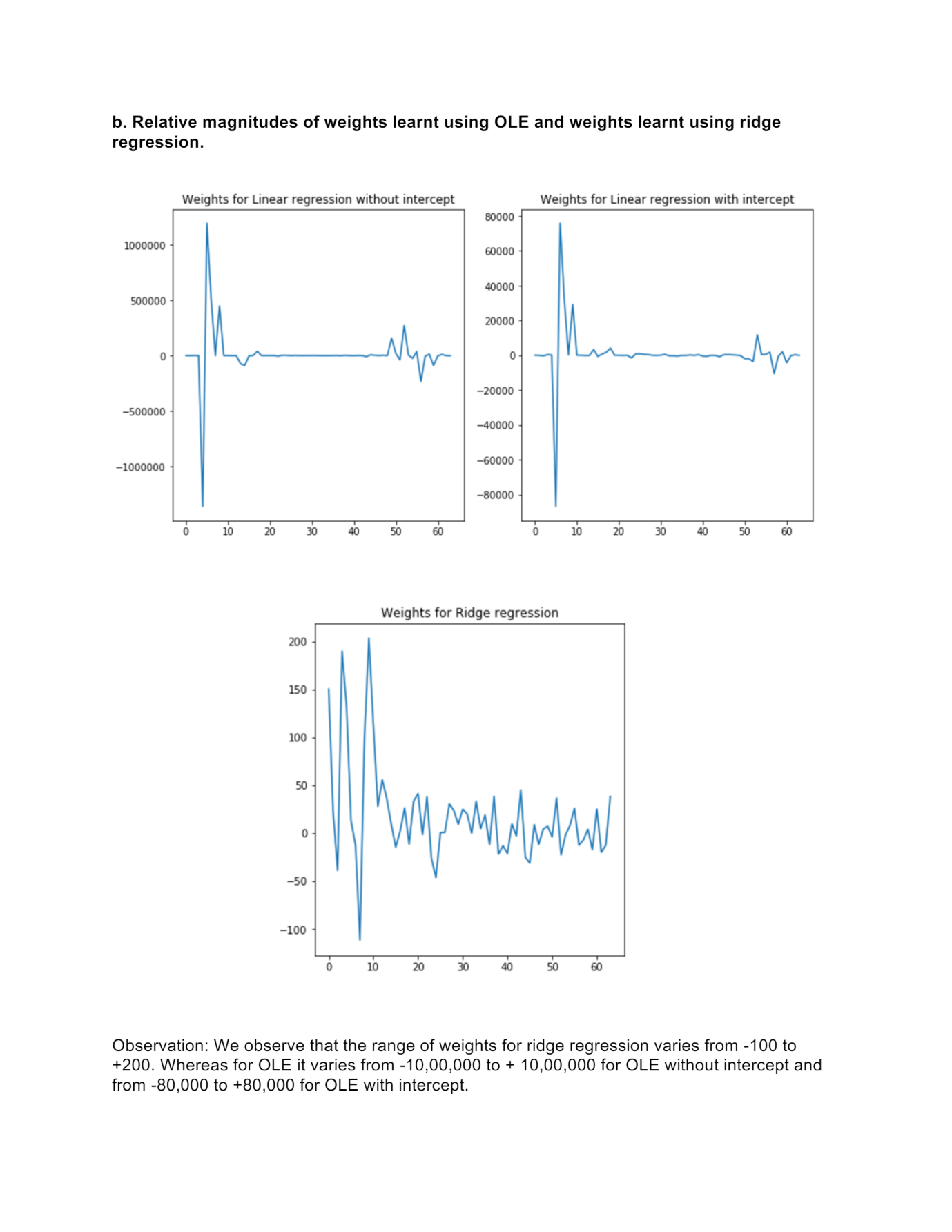


Figure 2

Table values for Mean Squared Error (MSE) at 0.06

|  |  |  |
| --- | --- | --- |
| Lambda | Test | Train |
| 0.0 | 3707.84 | 2187.16 |
| 0.01 | 2982.44 | 2306.83 |
| 0.02 | 2900.97 | 2354.07 |
| 0.03 | 2870.94 | 2386.78 |
| 0.04 | 2858.00 | 2412.11 |
| 0.05 | 2852.66 | 2433.17 |
| 0.06 | 2851.33 | 2451.52 |

Λ(lambda) = 0.06



|  |  |  |
| --- | --- | --- |
|  | Train Value | Test Value |
| OLE with intercept | 2187.16 | 3707.84 |

OLE intercept is similar with ridge regression. If we set optimal value at 0.06 we get

test value and training data at 0.0 since at that point the MSE is the least

**REPORT 4: Experiment with Gradient Descent for Ridge Regression Learning**

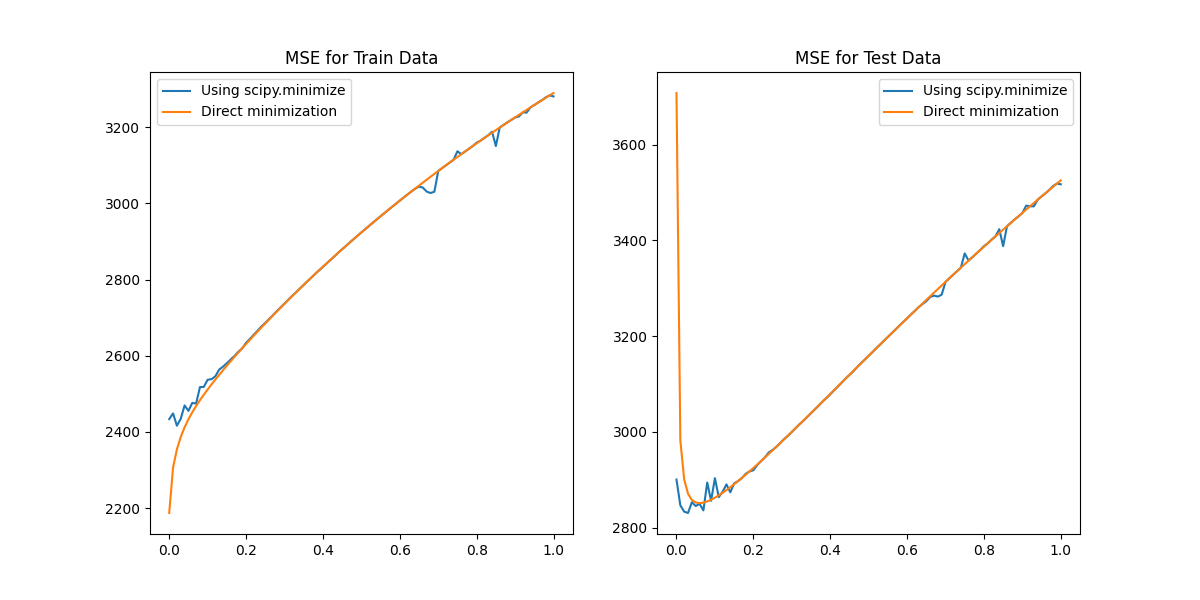
****

Figure 3:Gradient Descent for MSE

The above plotting for iteration 100.As we increase number of iterations the scipy.minimize learns more and gives results very similar to the direct ones.

|  |  |  |
| --- | --- | --- |
|  | Test | Train |
| MSE | 2451.5 | 2851.3 |

**REPORT 5: Non-Linear Regression**

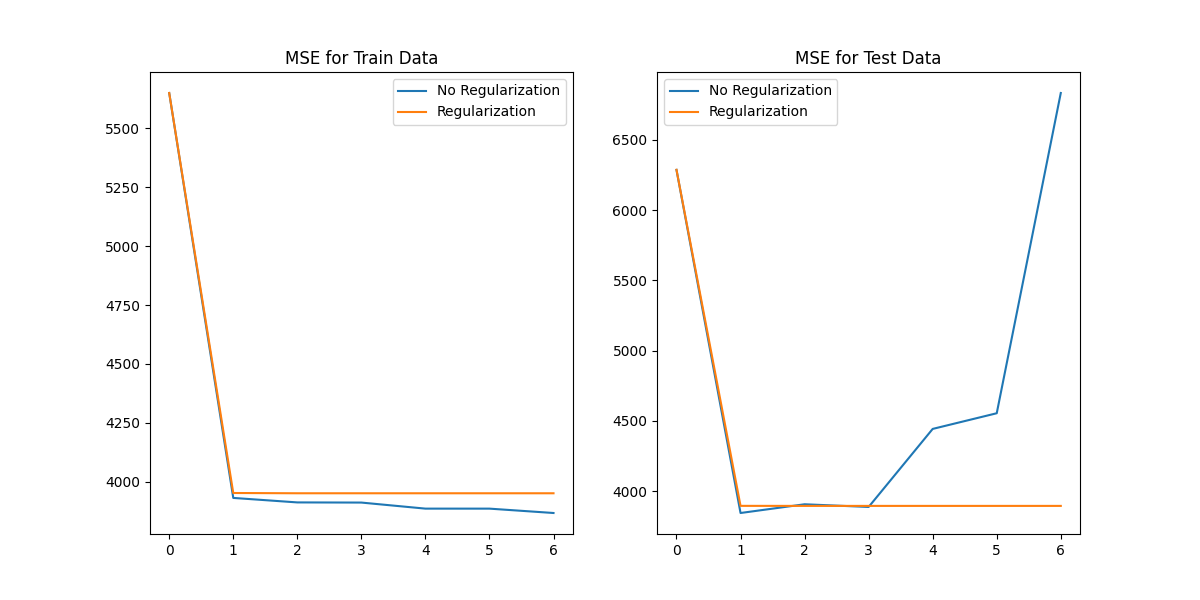
****

Figure 4:Non-linear regression

The optimal value of p 🡪 no regularization is 1

The optimal value of p 🡪 with regularization is 4

|  |  |  |
| --- | --- | --- |
|  | No regularization | With regularization |
| P 🡪1 | 3845.03 | 3895.85 |
| P 🡪4 | 4443.32 | 3895.58 |

**REPORT 6: Interpreting Results**

|  |  |  |
| --- | --- | --- |
| Model | Test | Train |
| Linear Regression - with intercept | 3707.84 | 2187.16 |
| Linear Regression - without intercept | 106775.36150046 | 19099.44684457 |
| Ridge Regression | 2451.53 | 2851.33 |
| Gradient descent ridge regression | 2451.53 | 2851.33 |
| Non-linear regression at p=1 | 3845.03473017 | 3895.85646447 |
| Non-linear regression at p=4 | 4443.32789181 | 3895.58266828 |

the best metric that can be choosen is ridge regression because ridge regression has lowest MSE values for test value and Train data.