CSE 574 INTRODUCTION TO MACHINE LEARNING

PROGRAMMING ASSIGNMENT 2

CLASSIFICATION AND REGRESSION

Team 8

Deepti Chavan (deeptisu)

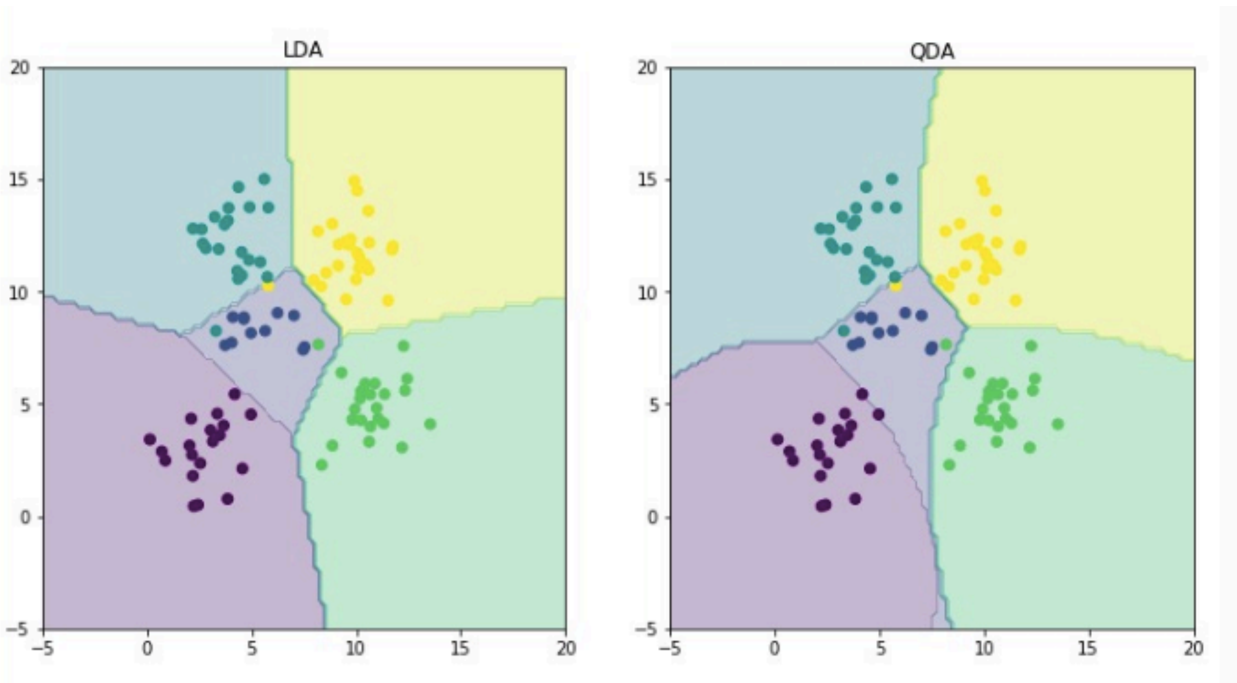
Kamalakannan Kumar (kkumar2)

Sushmita Sinha (ssinha7)

TABLE OF CONTENTS

PROBLEM 1: EXPERIMENT WITH GAUSSIAN DISCRIMINATORS

As the name suggests LDA is a linear classifier, it has a single covariance matrix (dxd) given a training data. Whereas, QDA is quadratic and has a covariance matrix for each output class. LDA assumes the same Σ for all classes, QDA computes a different Σ for each class. Essentially, LDA computes a separate μ for each class(using training points that belonged to it), but Σ is computed using the entire training data. *This is the reason for the skewed lines classifying the data set. In turn, the different approach to calculate covariance matrix for the data set explains the difference in two plots.*



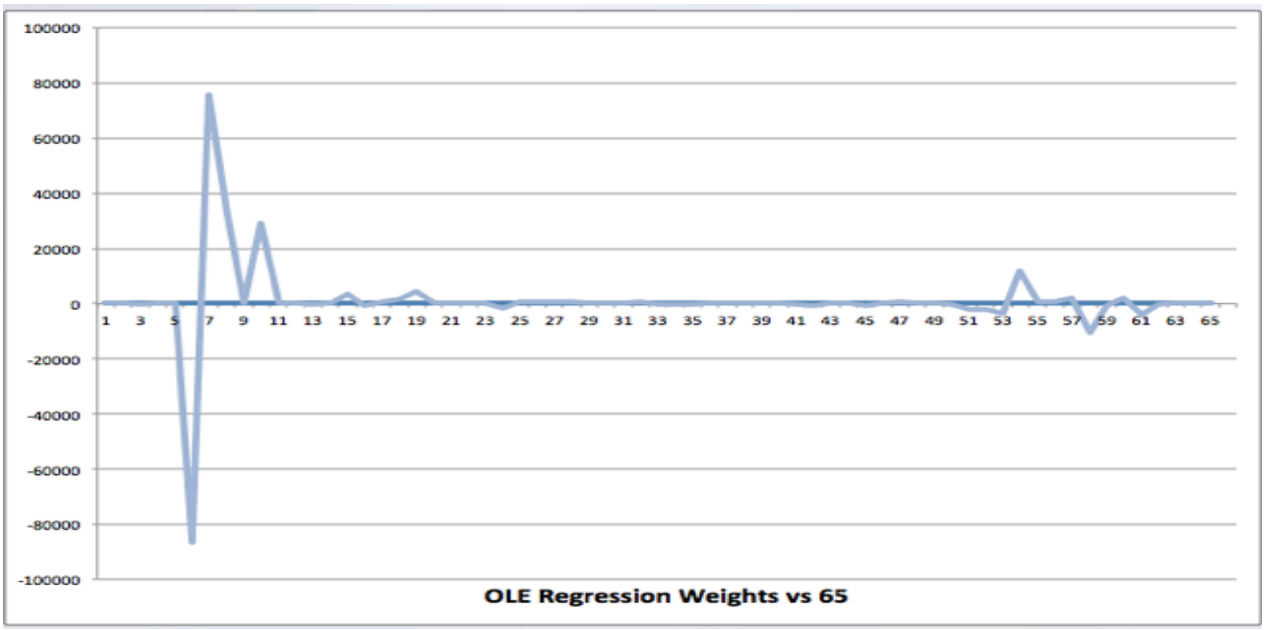
PROBLEM 2: EXPERIMENT WITH LINEAR REGRESSION

Problem 2	MSE without intercept	MSE with intercept
Training Data	19099.4468446	2187.16029493
Testing Data	106775.361558	3707.84018132

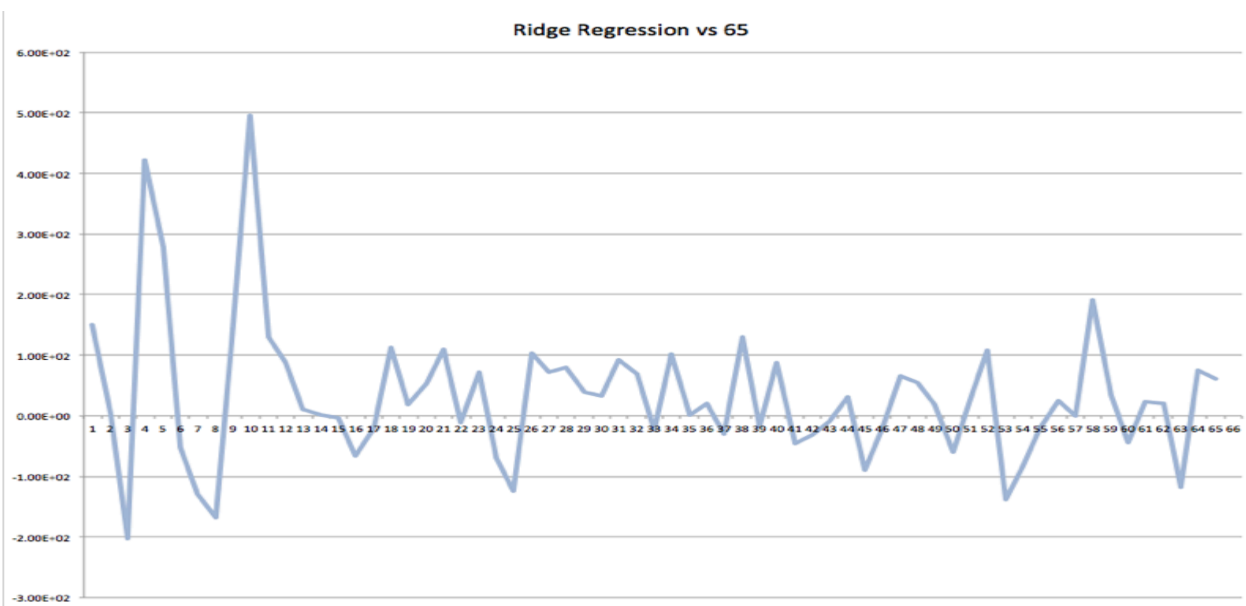
When we compute the MSE without an intercept (basically bias), we assume that the line equation will pass through the origin and this in-turn results in a higher MSE value. Addition of the bias value brings us closer to finding the line which fits the points in the concept area and hence gives a better result. Bias is the difference between the value estimate and the true value that should be obtained. Hence we get a better result and smaller MSE value with intercept present.

PROBLEM 3: EXPERIMENT WITH RIDGE REGRESSION

Problem 3	Lambda	MSE
Training Data	0	2187.16029493
Testing Data	0.06	2851.33021344

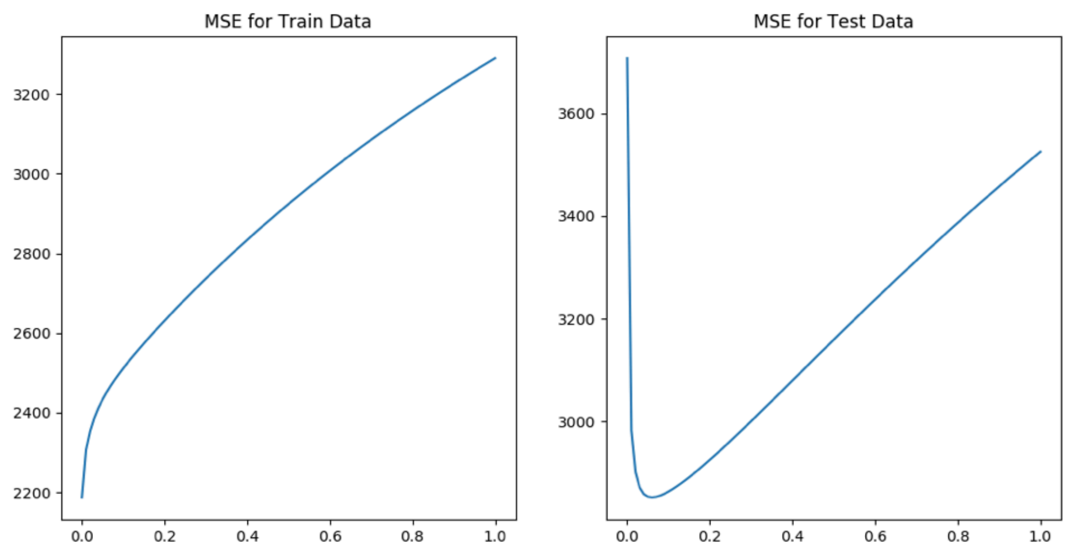


Relative magnitudes of weights learnt using OLE



Relative magnitudes of weights learnt using Ridge Regression

When we use parameter λ = 0, we can observe that MSE for linear regression and ridge regression are same over training data. When we are tuning the parameter λ, we obtain lesser MSE (Problem 3, λ = 0.06, MSE= 2851.33021344) value as compared to MSE in Problem 2 with no regularization (Problem 2, MSE =3707.84018132).
The regularization parameter λ is used to control the variance of the function. It is the shrinkage parameter and we can control the spread using λ. We need to choose a λ such that co-efficient are not rapidly changing. Variance of the ridge estimator vanishes as λ tends to infinity and variance of the ridge regression coefficient estimates decreases towards zero.



Graph for Problem 3: X-Axis: Lambda value v/s Y-Axis: MSE

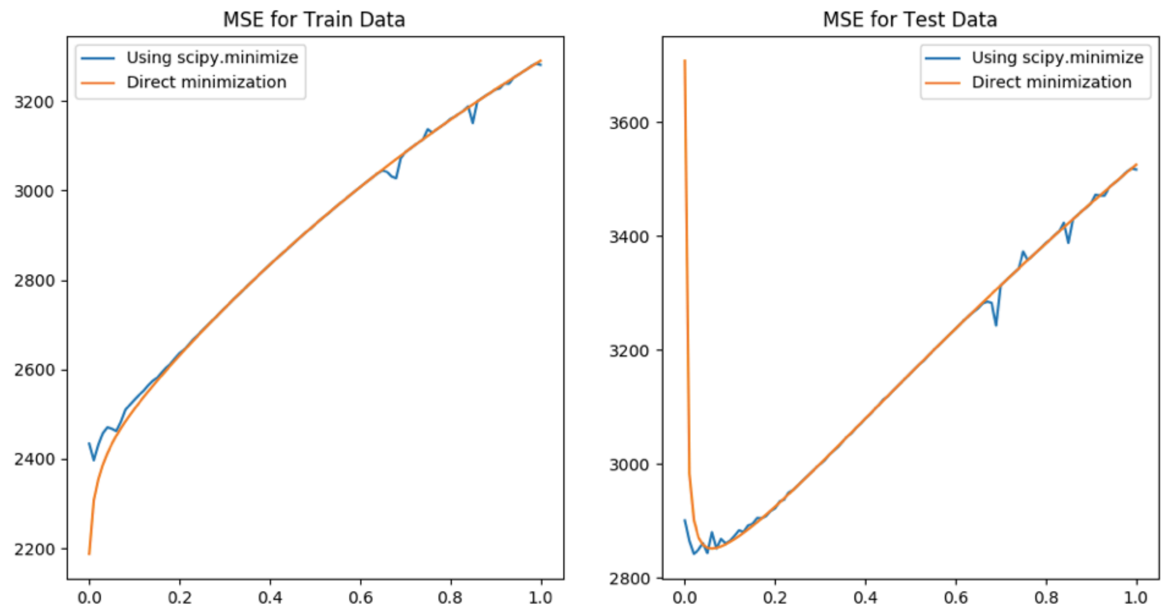
PROBLEM 4: USING GRADIENT DESCENT FOR RIDGE REGRESSION LEARNING

When comparing the MSE values obtained in Problem 3 and Problem 4, we observe that we can obtain the MSE value obtained in Problem 3 by increasing the number of iteration and obtain nearly same MSE values. As the number of iteration increases, we see that scipy.minimize and direct minimize overlay each other with a minimum number of outliners present between them.

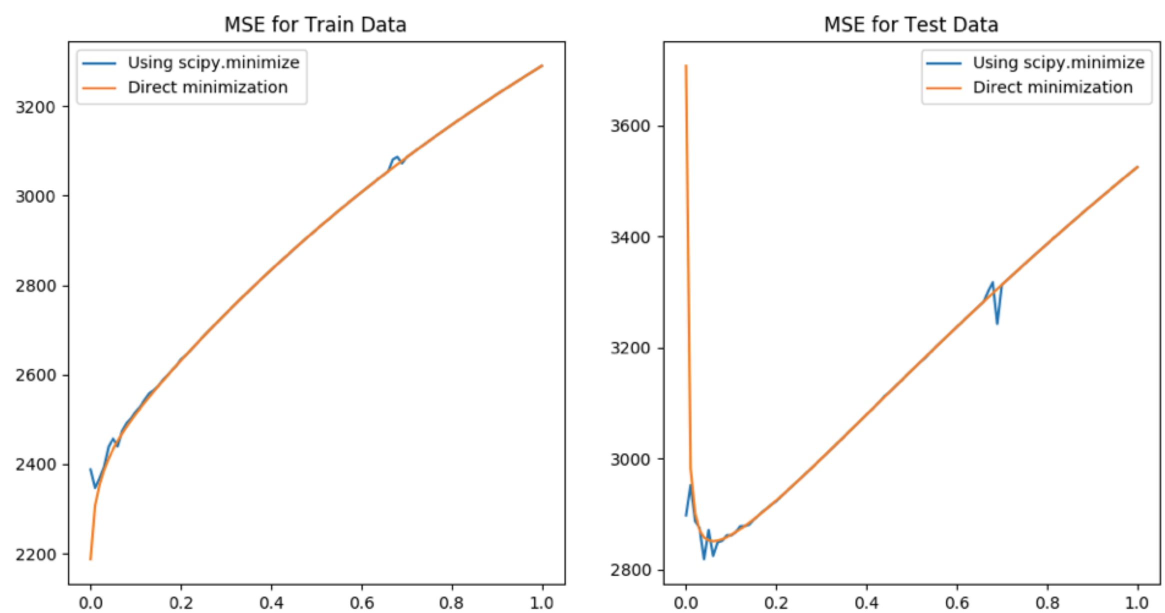
Problem 3	Lambda	MSE
Training Data	0	2187.16029493
Testing Data	0.06	2851.33021344

P-4, Training Data	iteration	MSE for λ = 0
1	20	2433.66541219
2	30	2387.91218901
3	40	2333.42382679
4	50	2309.77656757
5	100	2246.68872304

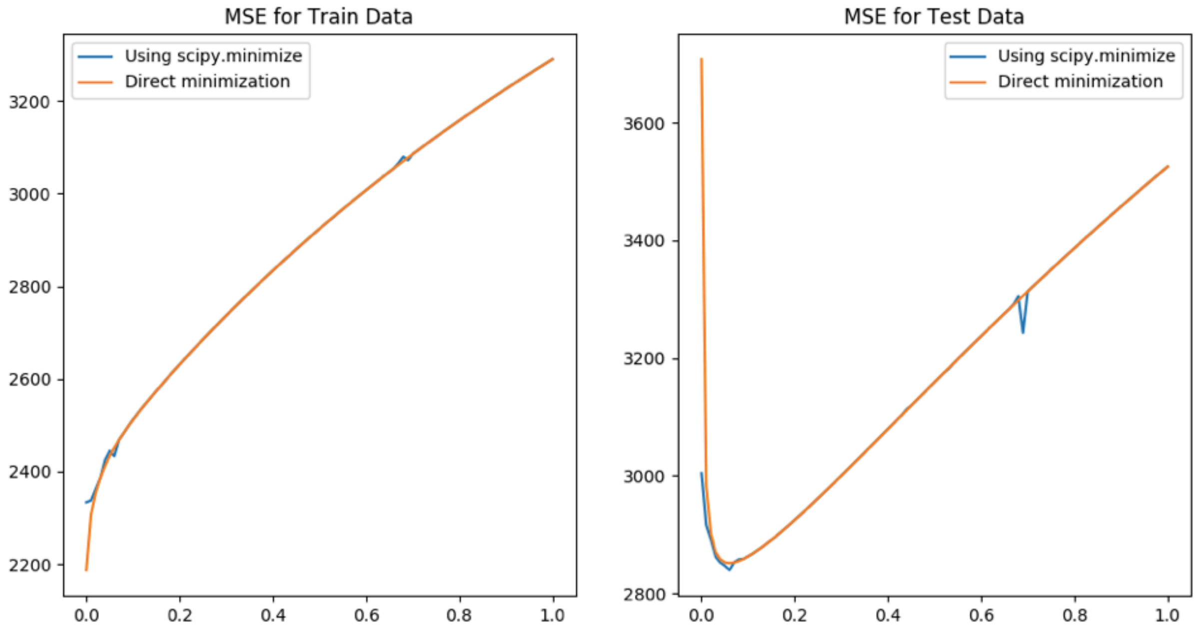
P-4, Testing Data	iteration	MSE for λ = 0.06
1	20	2879.73843814
2	30	2824.99914316
3	40	2839.6143869
4	50	2852.31302882
5	100	2851.45931557



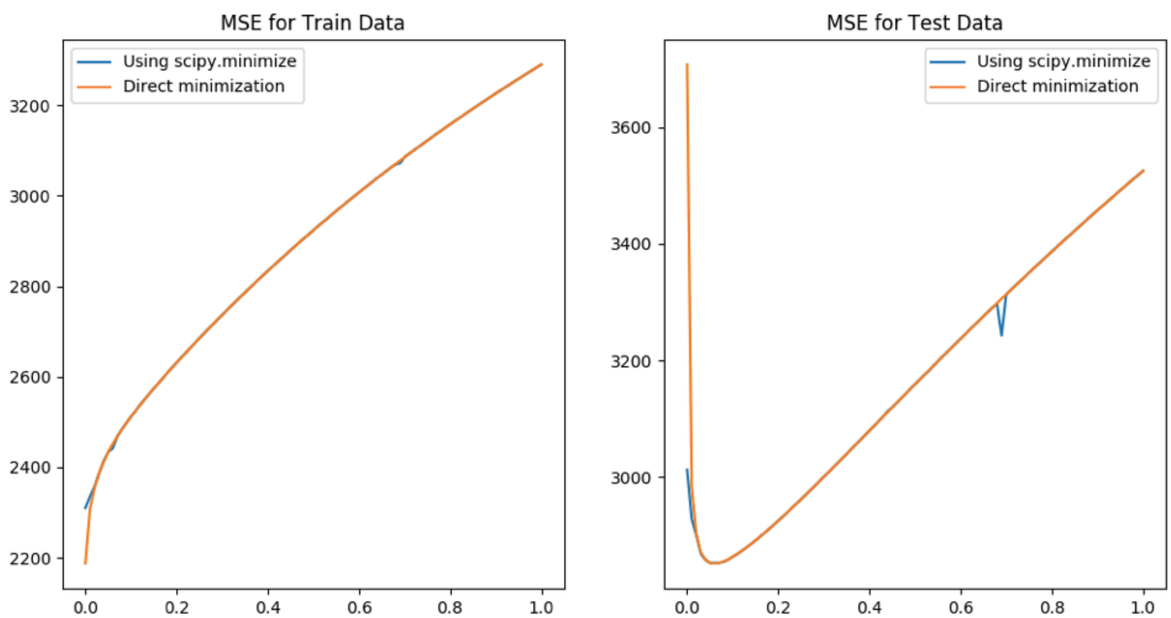
Graph for Problem 4: X-Axis: Lambda value v/s Y-Axis: MSE (iter = 20)



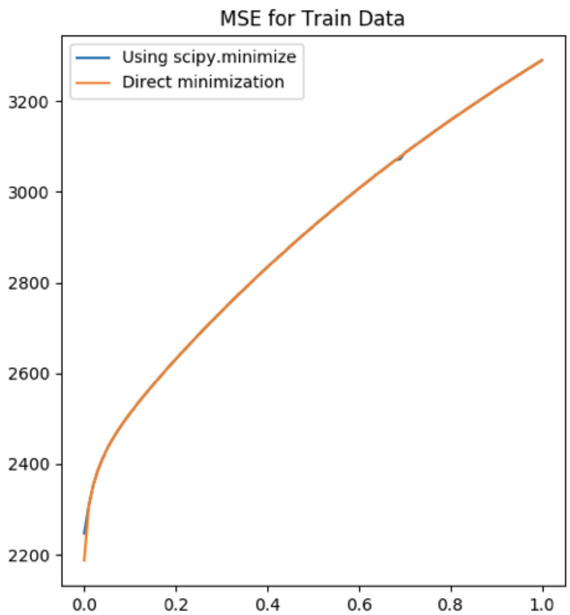
Graph for Problem 4: X-Axis: Lambda value v/s Y-Axis: MSE (iter = 30)

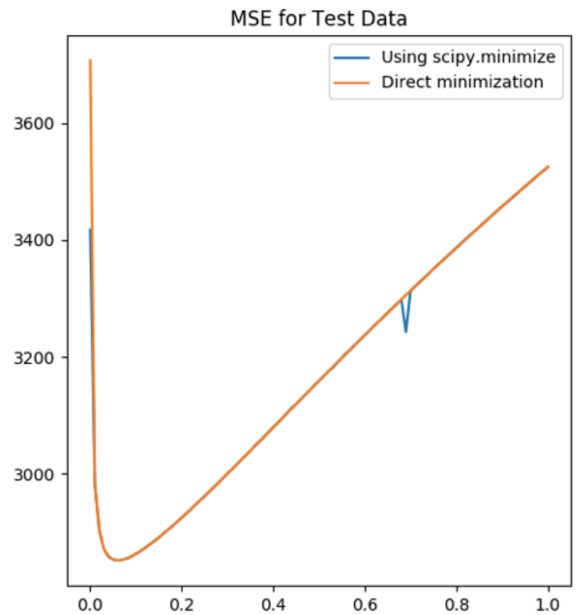


Graph for Problem 4: X-Axis: Lambda value v/s Y-Axis: MSE (iter = 40)



Graph for Problem 4: X-Axis: Lambda value v/s Y-Axis: MSE (iter = 50)





Graph for Problem 4: X-Axis: Lambda value v/s Y-Axis : MSE (iter = 100)

PROBLEM 5: NON-LINEAR REGRESSION

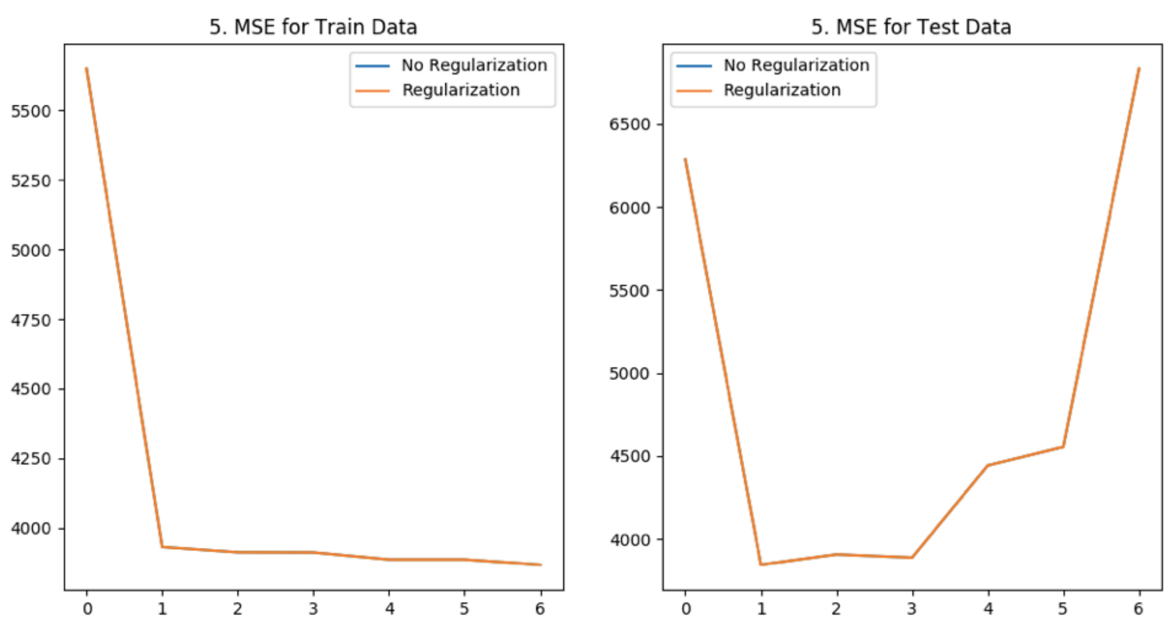
Using the λ = 0 and the optimal value of λ found in Problem 3 and varying p from 0 to 6, we obtain the following results.

P	P	MSE Training Data (λ = 0)
0	0	5650.7105389
1	1	3930.91540732
2	2	3911.8396712
3	3	3911.18866493
4	4	3885.47306811
5	5	3885.4071574
6	**6**	**3866.88344945**

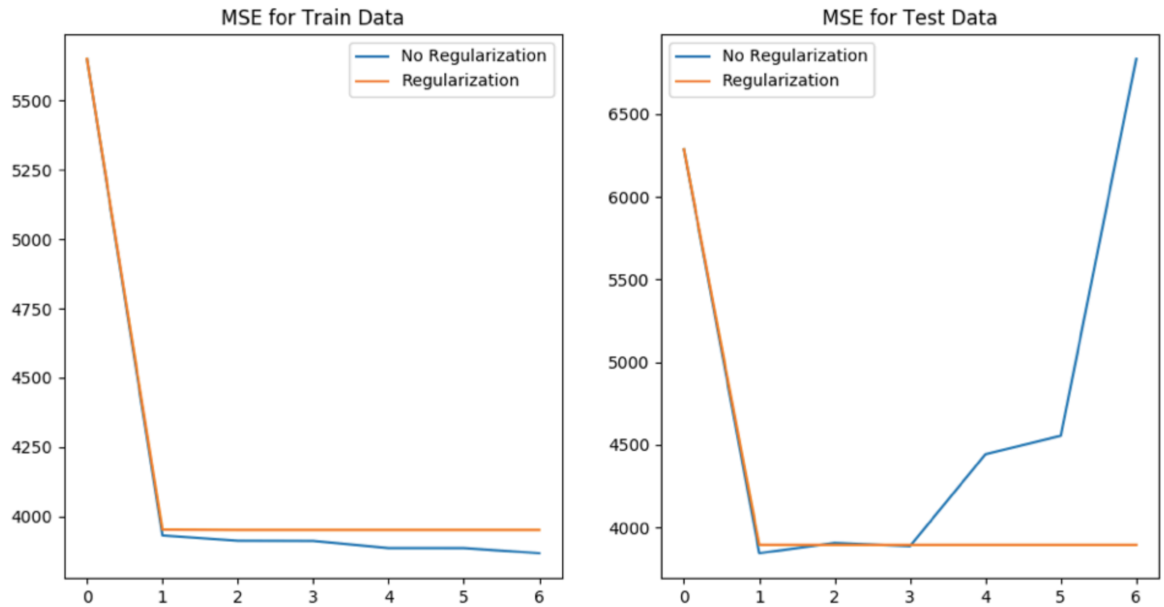
P	P	MSE Testing Data (λ = 0)
0	0	6286.40479168
1	**1**	**3845.03473017**
2	2	3907.12809911
3	3	3887.97553824
4	4	4443.32789181
5	5	4554.83037743
6	6	6833.45914872

Training Data(λ = 0.06)	P	MSE
1	0	5650.71190703
2	1	3951.83912356
3	2	3950.68731238
4	3	3950.68253152
5	4	3885.6823368
6	5	3880.68233518
7	**6**	**3866.88344945**

Testing Data (λ = 0.06)	P	MSE
1	0	6286.88196694
2	1	3895.85646447
3	2	3895.58405594
4	3	3895.58271592
5	**4**	**3895.58266828**
6	5	3895.58266872
7	**6**	3895.5826687



Graph for Problem 5: X-Axis: p-value v/s Y-Axis: MSE(λ = 0.0)



Graph for Problem 5: X-Axis: p-value v/s Y-Axis: MSE(λ = 0.06)

PROBLEM 6: INTERPRETING RESULTS

Summary of results obtained from the various methods are as follows:

The least MSE is obtained using ***Ridge Regression*** *having* ***MSE = 2851.33021344*** *and* ***λ = 0.06***

Problem 2	MSE without intercept	MSE with intercept
Linear Regression	106775.361558	3707.84018132

Problem 3	Lambda	MSE
Ridge Regression	0.06	2851.33021344

Problem 4	iteration	MSE for λ = 0.06
Gradient Descent for Ridge Regression	100	2851.45931557

Problem 5	P	MSE for λ = 0.06
Non-linear Regression	4	3895.58266828