PROGRAMMING ASSIGNMENT 1 – CLASSIFICATION AND REGRESSION

CSE 574 Fall 2021
Introduction to Machine Learning

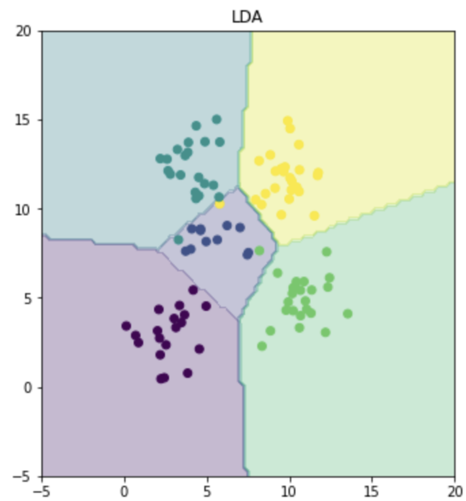
Group Number : 15

Submitted By :

S.No	Name	Email	UB Id Number
1	Lavanya Nadanasabapathi	lnadanas@buffalo.edu	50367936
2	Aravind Balakrishnan	balakri2@buffalo.edu	50412479
3	Anirudh Sowmyanarayanan	asowmyan@buffalo.edu	50415686

Problem 1 : Experiment with Gaussian Discriminators

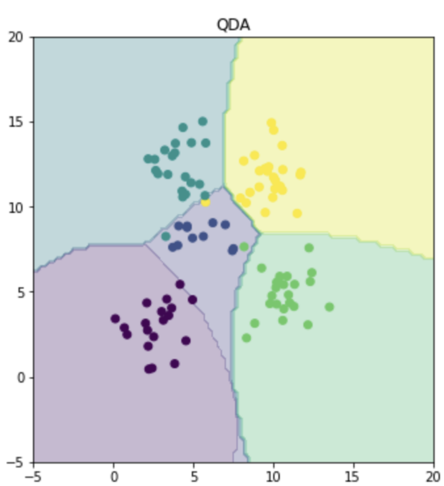
Linear Discriminant Analysis for our Training Data



1. The Accuracy on calculating Linear Discriminant Analysis on the given sample training data set is 0.97

2. LDA shows clear linear boundaries and better variance estimation as we have a small data set.

Quadratic Discriminant Analysis for our Training Data



1. The Accuracy on calculating Quadratic Discriminant Analysis on the given sample training data set is 0.96.

2. QDA doesn't show linear boundaries as they can be more accurate with large data and learns quadratic boundaries.

Comparing LDA and QDA with respect to our Training Data Set.

1. QDA has parabolic boundaries whereas LDA has Linear Boundaries.

2. LDA uses a lesser number of parameters compared to QDA as QDA has a separate covariant matrix for every class.

3. For small data sets, LDA gives better variance estimation than QDA.

To Summarize, In General, QDA has a better accuracy than LDA for larger data sets but with respect to our training model, Linear Discriminant Analysis has a better accuracy than Quadratic Discriminant Analysis, the accuracy respectively is 0.97 and 0.96. Which shows that LDA is better for our sample training data.

Problem 2. Experiment with Linear Regression

MSE Values :

	Training Data	Test Data
Without Intercept	19099.446844570688	106775.3615622583
With Intercept	2187.160294930391	3707.840181218663

The above table has results which shows that the MSE without intercept is much greater than when using intercept for both the test and training samples.

Statistics :

1. Training Data with With Intercept is almost 8.7 times less than without an intercept.

2. Test Data with Intercept is almost 28.8 times less than without an intercept.

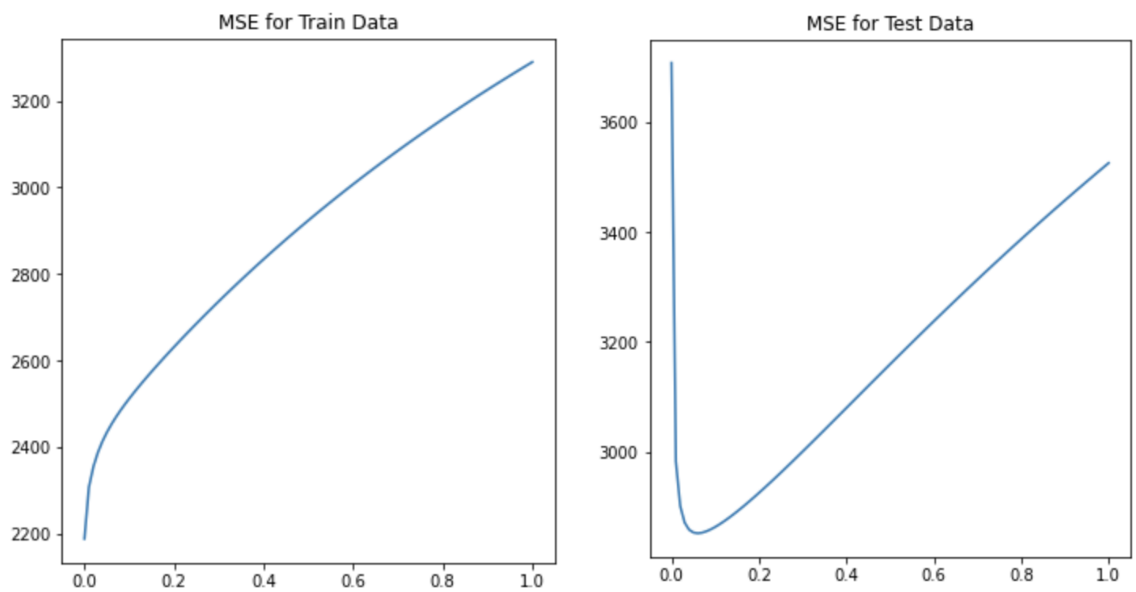
3. With intercept we observe that there is a considerable decrease in error for both training and test data.

Problem 3 : Ridge Regression

The MSE for Ridge Regression with Intercept :

Training Data : 2187.16029493
Test Data : 2851.33021344

<u>Plot MSE for Training and Testing Data :</u>



<u>Comparison of Mean Weights of Ridge Regression and Linear Regression</u>

OLE weight : 3707.84
RIDGE weight : 2851.33

On comparing the mean weights, It can be understood that the Ridge regression is better than OLE as it has a lower mean weight.

<u>Comparison of Error for Ridge Regression and Linear Regression</u>

Model	**Testing Data**	**Training Data**
OLE	3707.84	2187.16
Ridge Regression	2851.33	2187.16

On comparing the error values, we see the error for training data
Calculation of Lambda :

The value of lambda ranges from 0 to 1, increasing in steps by 0.01.
The table below represents the first 20 values of lambda :

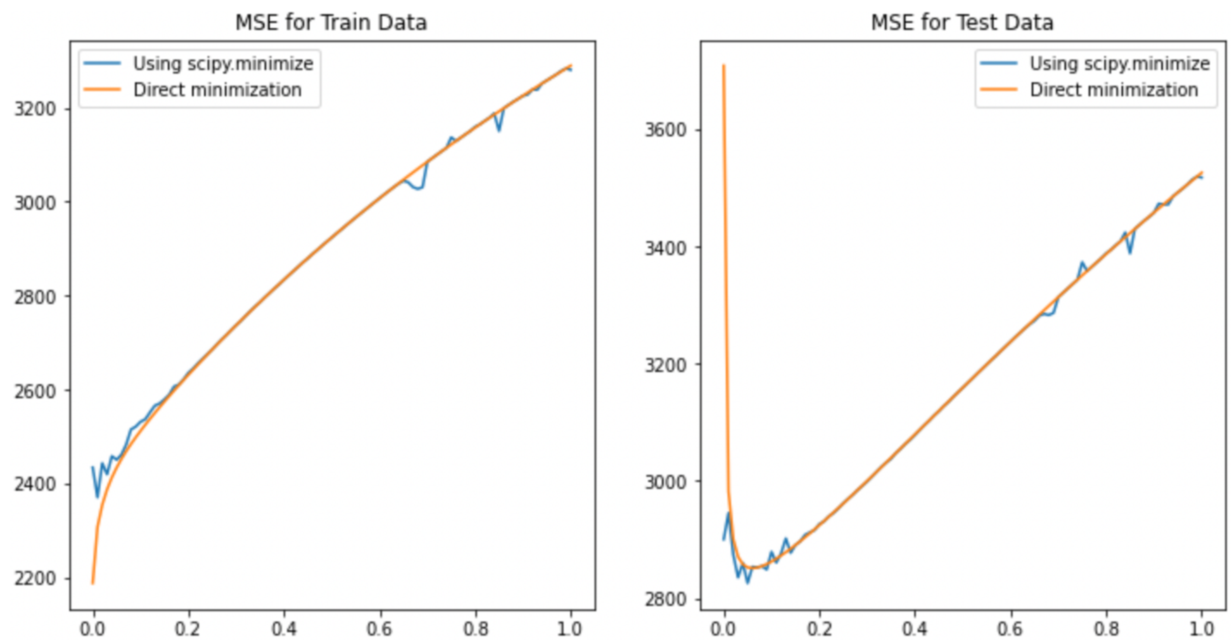
Lambda (λ)	**Train Data**	**Test Data**
0.00	2187.16029493	3707.84018122
0.01	2306.83221793	2982.44611971
0.02	2354.07134393	2900.97358708
0.03	2386.7801631	2870.94158888
0.04	2412.119043	2858.00040957
0.05	2433.1744367	2852.66573517
0.06	2451.52849064	2851.33021344
0.07	2468.07755253	2852.34999406
0.08	2483.36564653	2854.87973918
0.09	2497.74025857	2858.44442115
0.10	2511.43228199	2862.75794143
0.11	2524.60003852	2867.63790917
0.12	2537.35489985	2872.96228271

0.13	2549.77688678	2878.64586939
0.14	2561.92452773	2884.62691417
0.15	2573.84128774	2890.85910969
0.16	2585.55987497	2897.30665895
0.17	2597.10519217	2903.94112629
0.18	2608.49640025	2910.73937213
0.19	2619.74838623	2917.68216413

The above table suggests the MSE value for testing and training data where λ = 0.06 is the lowest testing error. After this conjecture, the MSE value starts increasing with the increasing value of lambda (λ) .

Problem 4 : Ridge Regression using Gradient Descent

The Plot for MSE vs Lambda for testing and training data for Ridge Regression using Gradient Descent :

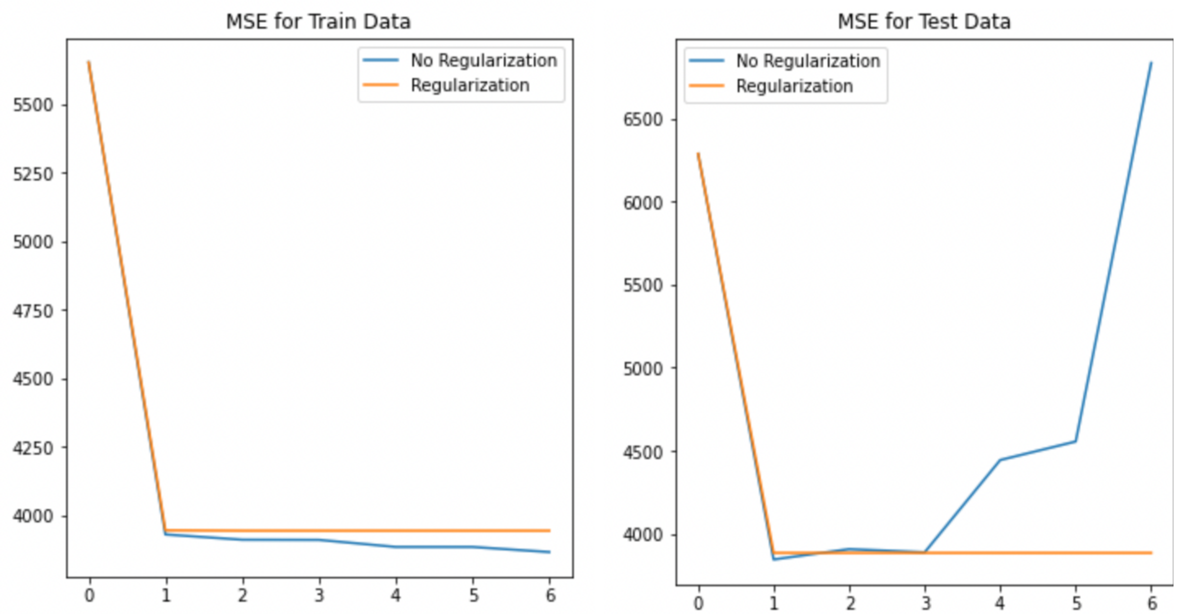


The lowest value of MSE using gradient descent is **2187.16130594.** Comparing the values observed from Ridge Regression is 2187.16029493 (Problem 3), the values are almost similar.

The Lamda**(λ)** observed in both problems are the same 0.06.

Problem 5 : Non-linear Regression

Plot for Train and Test data :



Here, both the errors are initially high, then flattened after 1. Then the Zero regularization increases again.

Values for train data:

P	**MSE 0 Regularization**	**MSE With Regularization**
0	5650.7105389	5650.71148907
1	3930.91540732	3945.99483429
2	3911.8396712	3944.6776809
3	3911.18866493	3944.67305264
4	3885.47306811	3944.67283303
5	3885.4071574	3944.67283129
6	3866.88344945	3944.67283125

Values for test data:

P	MSE 0 Regularization	MSE with Regularization
0	6286.40479168	6286.80226417
1	3845.03473017	3884.6956229
2	3907.12809911	3884.54567903
3	3887.97553824	3884.54664192
4	4443.32789181	3884.54663566
5	4554.83037743	3884.5466369
6	6833.45914872	3884.54663693

It is observed from the above graph and the table that the test error with optimal P is 1 for zero value of Lambda and P is 4 for optimal value of Lambda.

When regularisation isn't used, after it reaches its lowest where P = 1, it steadily increases till P = 5 and then rapidly increases between P = 5 and P = 6.

This model with increasing higher order polynomials will face overfitting; the training error decreases for the given data. However, for a completely new data set the error will steeply increase.

When regularisation is used, the error decreases to a minimum at P = 5. But post that the error value will not increase drastically.

This model with increasing higher order polynomials the weights will be low and result in a smooth curve, which works better in any general data set.

Problem 6 : Interpreting results

The results obtained using the various regression techniques as summarized as belows:

<u>OLE Regression</u>

MSE for Test Data without Intercept	106775.3615622583
MSE for Test Data with Intercept	3707.840181218663
MSE for Training Data without Intercept	19099.446844570688
MSE for Training Data with Intercept	2187.160294930391

<u>Ridge Regression</u>

Optimal $\lambda = 0.06$

Training Error : 2451.52849064
Testing Error : 2851.33021344

<u>Ridge Regression with Gradient Descent</u>

Optimal $\lambda = 0.06$
Testing Error : 2187.16130594

<u>Non Linear Regression</u>

Training Error with regularization : 3944.67283125
Training Error without regularization : 3885.4071574

Testing Error with regularization : 3884.54663566
Testing Error without regularization : 3887.97553824

Conclusion :

On comparing the errors obtained from various methods we can say that Ridge Regression and Gradient Ridge Regression have the lowest test error and hence the best approach while performing linear regression on the trained data.