University at Buffalo
CSE 574 Introduction to Machine Learning

Programming Assignment 1 Report

Submitted by: **Group 95**
Vishal Chaurasia – 50353581
Pushkar K Pandey – 50412529
Varun Sudarshan – 50360324

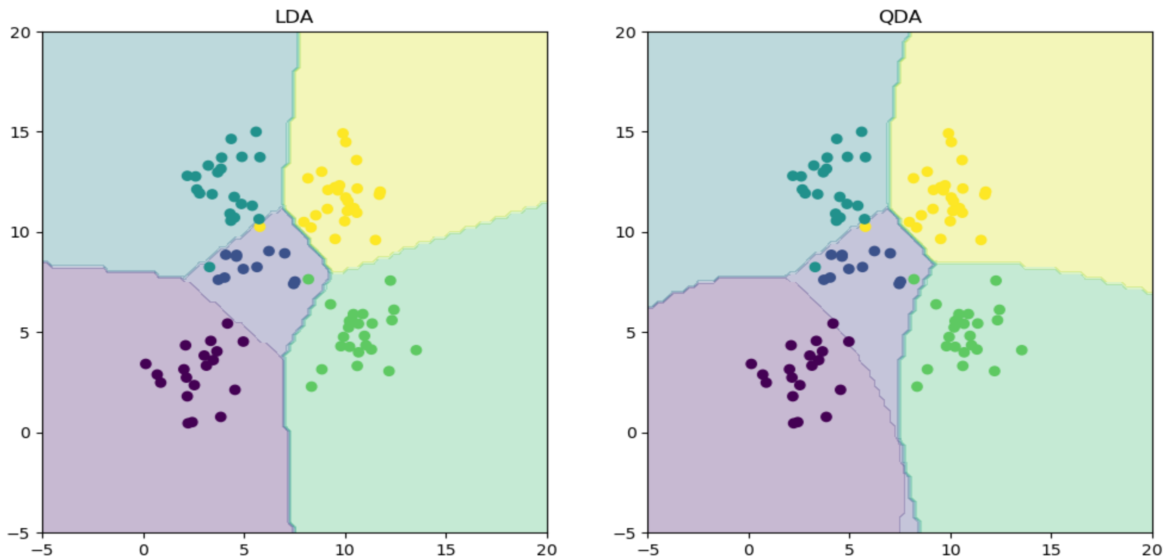
October 12, 2021

<u>Problem 1: Experiment with Gaussian Discriminators</u>

LDA Accuracy: 97%
QDA Accuracy: 96%

As seen from the accuracy data of the LDA & QDA the accuracy of LDA is better than the QDA this probably because of the small sample size. Additionally, from the graph depicted below we can see below the two data point in the purple region are far inside the boundary but in case of QDA the data points lie exactly at the boundary between the purple & violet region probably this may be the reason behind 1% lower accuracy of QDA than that of LDA.

Discriminating Boundaries for LDA and QDA is as below:



From the above two graphs we can see that the boundaries for the LDA graph are linear, while that of the QDA is non-linear. This is primarily because, in LDA we consider the same covariance matrix for the entire data irrespective of the different labels. On the other hand, in case of QDA different covariance matrices are computed and maintained for each of the unique labels in the data set.

Problem 2: Experiment with Linear Regression

Data Category	MSE	
	Without Intercept	With Intercept
Train	19099.446844570935	2187.1602949303897
Test	106775.36155426428	3707.840181595626

As seen above, the MSE for the training & test data is much lower when weights were calculated with intercept, therefore we can conclude that it is better to use intercept while computing weights.

Problem 3: Experiment with Ridge Regression

Used ridge regression parameters with intercept and the testOLERegression function to calculate the MSE for training and test data. The train and test data errors for different λ values (0 to 1 in steps of 0.01) are plotted below.

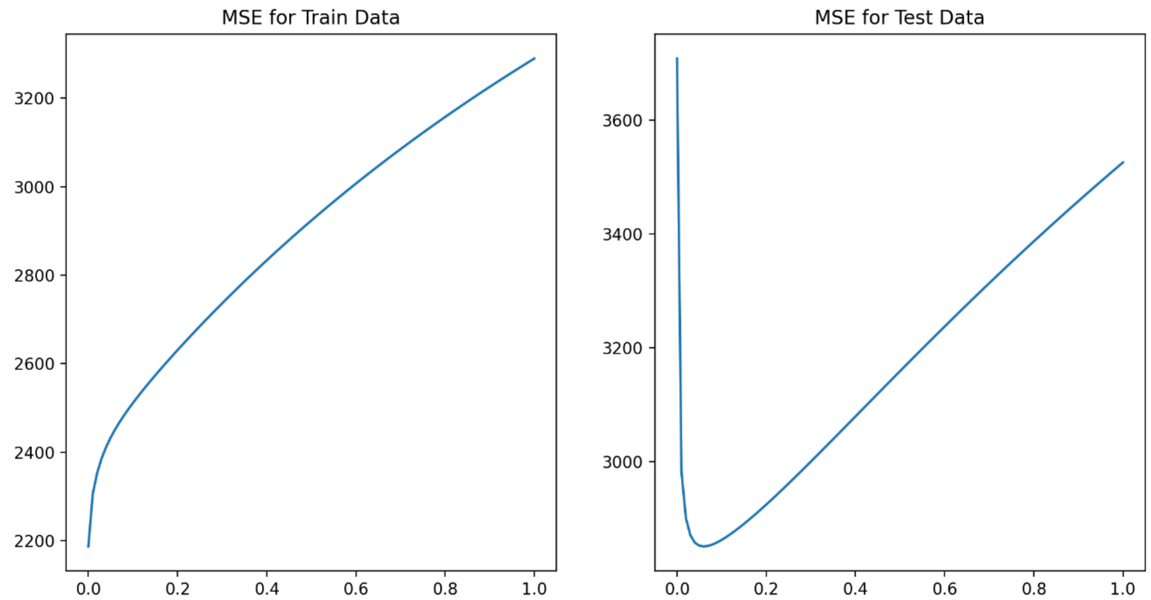


Fig: MSE for training and test data using ridge regression parameters

From the left plot of the MSE on the training dataset, we observed that MSE increases if the λ value increases, and we get the lowest MSE if λ=0. Thus, an increase in λ value indicates more penalty for learning, which is reflected by the high MSE value.

From the right plot of the MSE on the testing dataset, we observed that MSE is the lowest when λ = 0.06. However, once it crosses this value, the MSE increases, reflecting a poor fit for the testing dataset.

The optimal value for **λ is 0.06** in the case of test data and 0 in the case of training data because, at this point, we get the lowest value of MSE. p

Comparing between the relative magnitudes of weights learnt using OLE and weights learnt using ridge regression:

From the below table comparing values for ridge weights and OLE, we can state that Ridge regression is lower than the weights of OLE as we are regularizing the weights in Ridge regression.

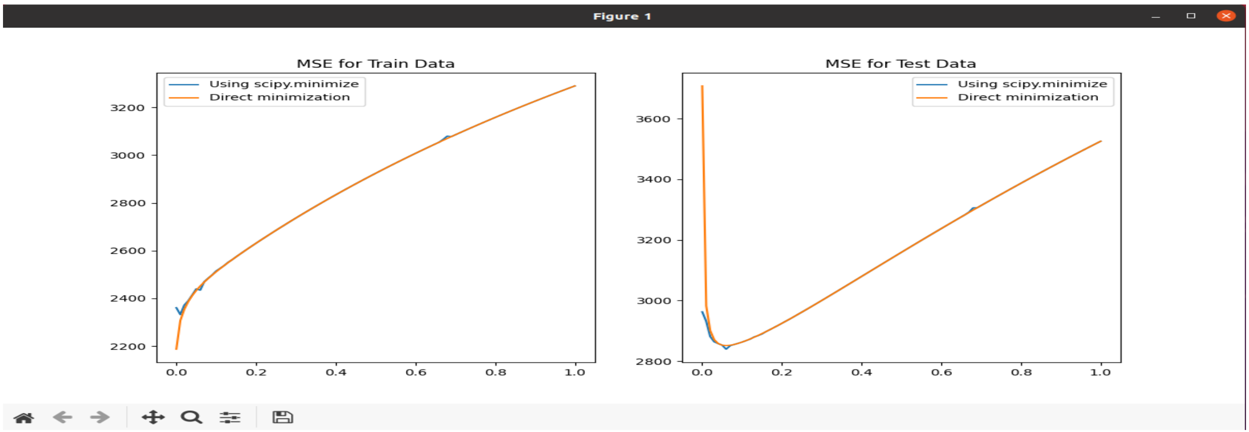
RIDGE WEIGHTS	OLE WEIGHTS
150.45959807	148.154876
4.80776899	1.27485192
-202.90611468	-293.38352206
421.7194576	414.72544846
279.45107288	272.08913421
-52.29708233	-86639.45715596
-128.59418907	75914.46805149
-167.50057028	32341.62283186
145.74068096	221.1012135
496.30604123	29299.55120569
129.94845775	125.23036028
88.30438076	94.4110833
11.29067689	-93.86286325
1.88532531	-33.72827994
-2.58364157	3353.19771053
-66.89445481	-621.0963091
-20.61939955	791.73653265
113.39301454	1767.76038918
17.99086827	4191.6740567
52.50235963	119.43812095
109.68765513	76.61034003
-10.72779629	-15.2001293
71.67974829	82.24245935
-69.30906366	-1456.6620857
….	….

Comparing the two approaches in terms of errors on train and test data

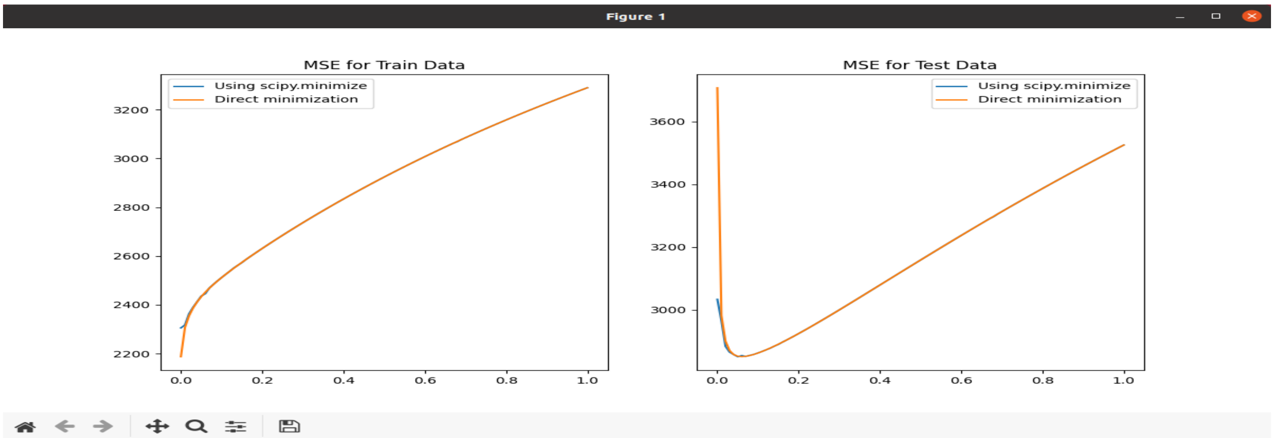
For OLE, when considering data with intercept, the MSE value is 3707.840181735625, but we get 2851.33021344 on the test data result with the Ridge regression method, which is a lower value. Based on this result, we can state that we get lower MSE for testing in Ridge regression.

Problem 4: Using Gradient Descent for Ridge Regression Learning

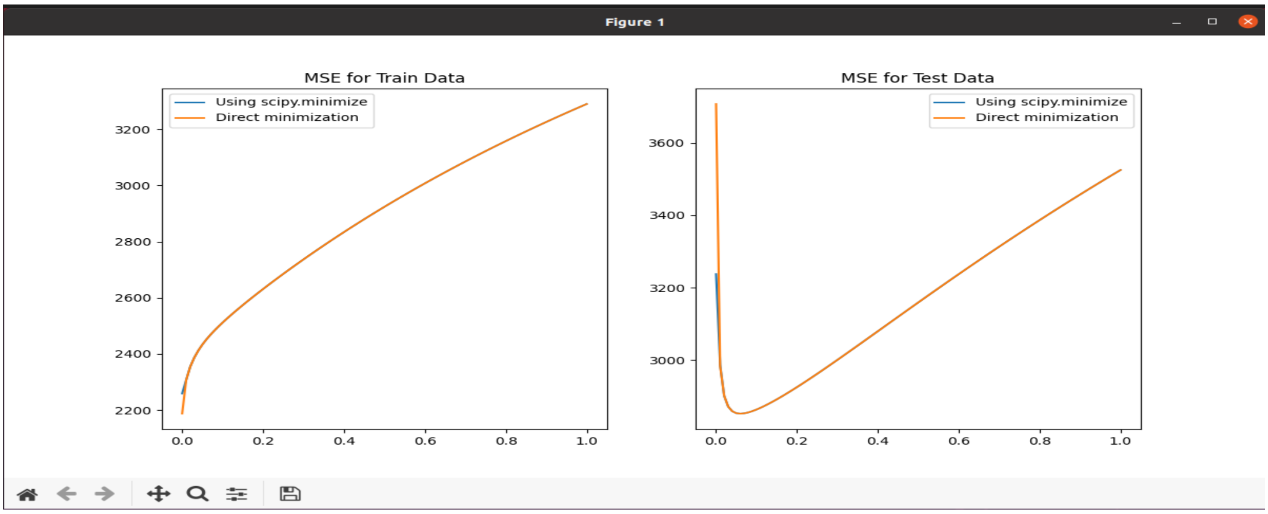
For iteration = 40



For iteration = 50



For iteration = 100



By the graph, we can tell that MSE is lower for gradient descent-based learning

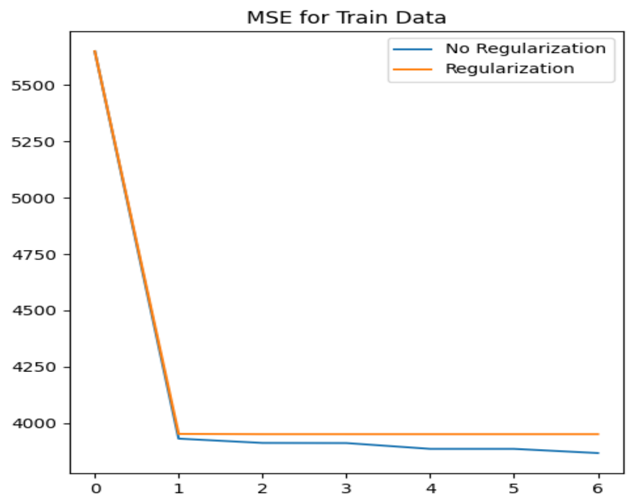
- MSE without Gradient Descent: 2851.33021344

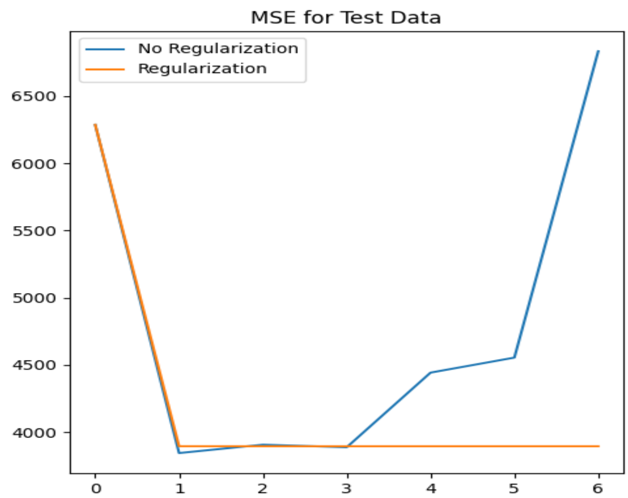
- MSE with Gradient Descent: 2833.13531834
- λ without Gradient Descent: 0.06
- λ with Gradient Descent: 0.02

Problem 5: Non-Linear Regression

p	0	1	2	3	4	5	6
MSE with $\lambda = 0$	6286.40479168	**3845.03473017**	3907.12809911	3887.97553824	4443.32789181	4554.83037743	6833.45914872
MSE with Optimal λ	6286.88196694	3895.85646447	3895.58405594	3895.58271592	**3895.58266828**	3895.5826687	3895.58266872

When λ = 0, Optimal value of p is 1
For optimal λ, Optimal value of p is 4





Problem 6: Interpreting Results

Model	MSE			
	Training Data		Test Data	
Linear OLE Regression with Intercept	2187.1602949303897		3707.840181595626	
Linear OLE Regression without Intercept	19099.446844570935		106775.36155426428	
Rigid Regression	2451.52849064		2851.33021344	
Rigid Regression with gradient descent	**2375.768565**		**2833.13531834**	
	λ= 0	**λ= 0.06**	**λ= 0**	**λ= 0.06**
Non-Linear Regressing with p=1	3930.91540732	3951.83912356	**3845.03473017**	3895.85646447
Non-Linear Regressing with p=4	3885.47306811	3950.6823368	4443.32789181	**3895.58266828**

Based on the above data collected, we recommend Ridge Regression with gradient descent to predict diabetes.

The recommendation is based on the following considerations:

1. We will achieve the best result when we use Gradient Descent for optimization with regularization parameter λ= 0.02.

2. Ridge Regression with gradient has the least MSE value (2833.13531834) on test data compared to all distinct models.

3. We can avoid high computation of $(X^{\top}X)^{-1}$ when we use Gradient Descent

4. Low MSE value indicates that the deviation from the true label is the least; therefore, the prediction accuracy will be high.