50291708
50291555
50291152

Programming Assignment 1

CSE 574
March 09, 2019

Problem 1

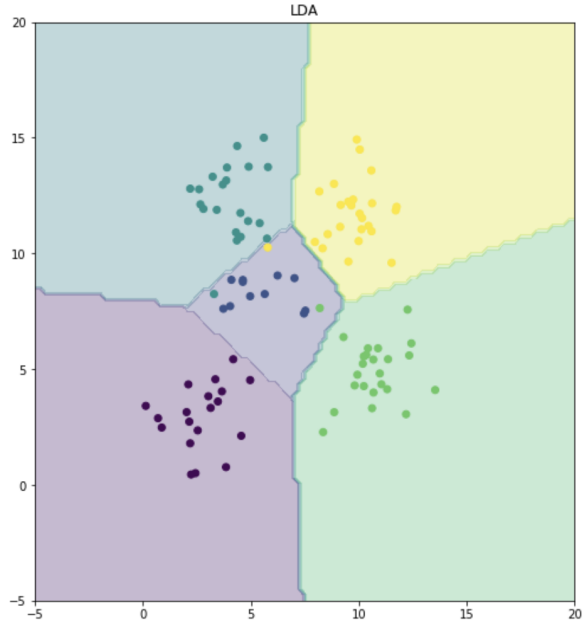
LDA Accuracy = 0.97
QDA Accuracy = 0.96

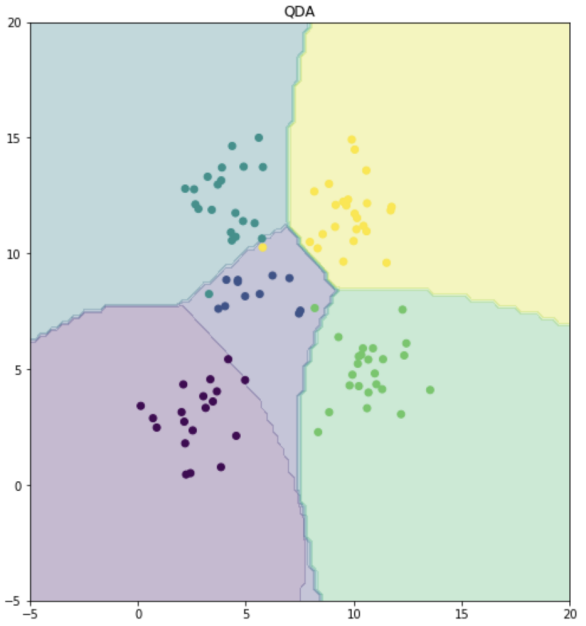
We see that LDA has a higher accuracy than QDA. This is because we have only 242 patient records for learning. In LDA, we are learning the covarinace matrix from all 242 records but in QDA we are learning covariance matrix for each labels. So, in QDA there is small number of records in each label. Due to very small amount of record in each label we get less accuracy in QDA.

The decision boundary is almost linear for LDA and non-linear for QDA. This is because LDA considers the whole dataset (only 1 label) while learning the covarince matrix. That's why the decision boundary is linear for LDA. But for QDA the boundary is non-linear as it considers all the unique labels in the dataset.

From the graph we see that there are 2 data points in QDA that is on the boundary line between purple and violet and becuase of these two data point we see the difference in the boundary line between purple and violet. Also may be these two data point are also responsible for 1 percent less accuracy in QDA.

Discriminating boundary for LDA and QDA is shown below :





Problem 2 : Linear Regression

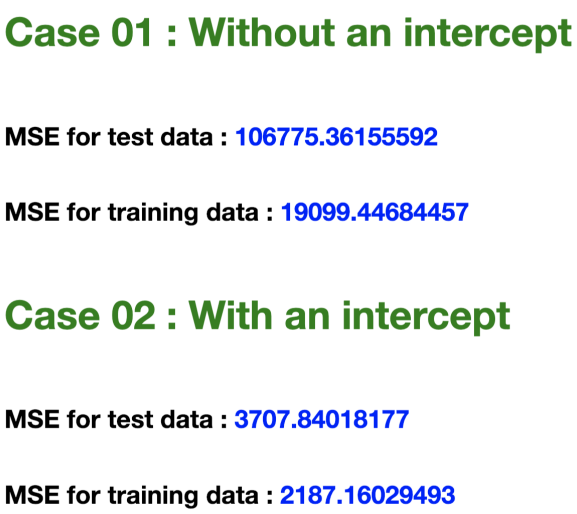
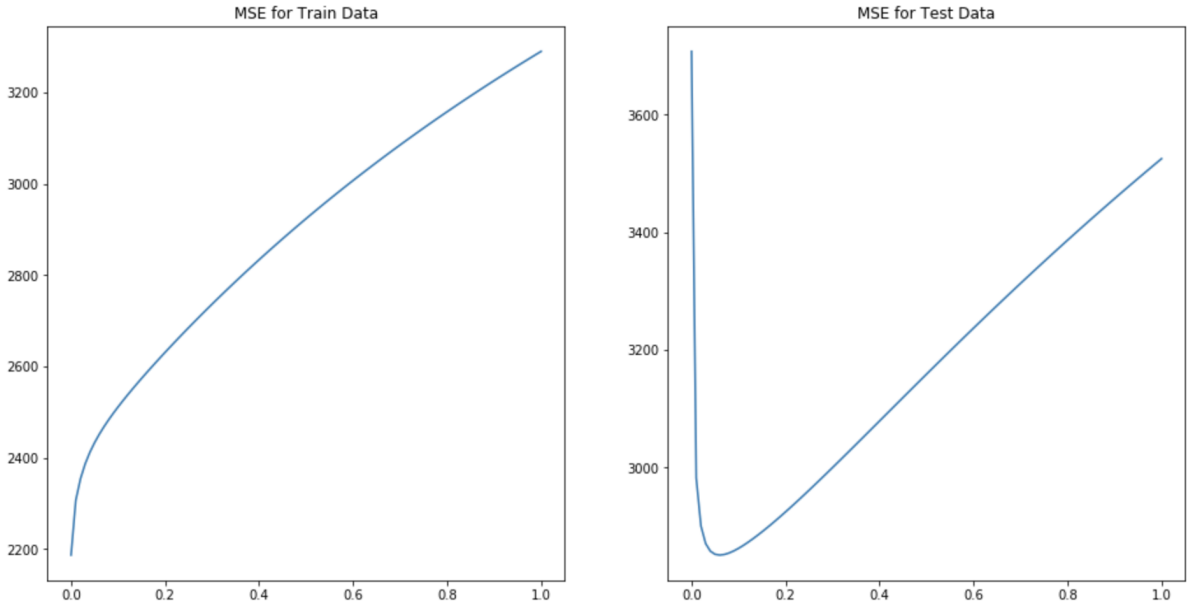


Figure 1: MSE for training and test data

MSE is lower for the second case (using intercept) both for training data and test data. So, for the given dataset it is better to use an intercept for learning the weight vector for regression. Also we can observe that when we are using the intercept then the change in MSE is higher for the test data.

Problem 3 : Ridge Regression

We calculated the MSE for training and test data using ridge regression parameters and the **testOLERegression** function. We also used the data with intercept. The errors on train and test data for different λ values (from 0 to 1 in steps of 0.01) are plotted below.



The left plot is the MSE for the train dataset. We observe that MSE is increasing if the λ value increases and we get the lowest MSE if λ=0 (no regularization). Increase in λ value indicates more penalty for learning which is reflected by the high MSE value.

The right plot is the MSE for the testing dataset. We find that MSE will be lowest when λ=0.06. After this value if we increase the λ value then MSE starts increasing which reflects poor fit for the testing dataset.

The optimal value for λ is **0.06**. Because in this value the MSE is lowest for testing data.

Comparison between OLE errors and Ridge regression errors:

We are considering data with intercept. In OLE method, MSE for test data is 3707.84018177 but in Ridge regression method, MSE for test data is 2851.330213443848. So, we get lower MSE for testing if we use Ridge regression.

Comparison between OLE weights and Ridge regression weights:

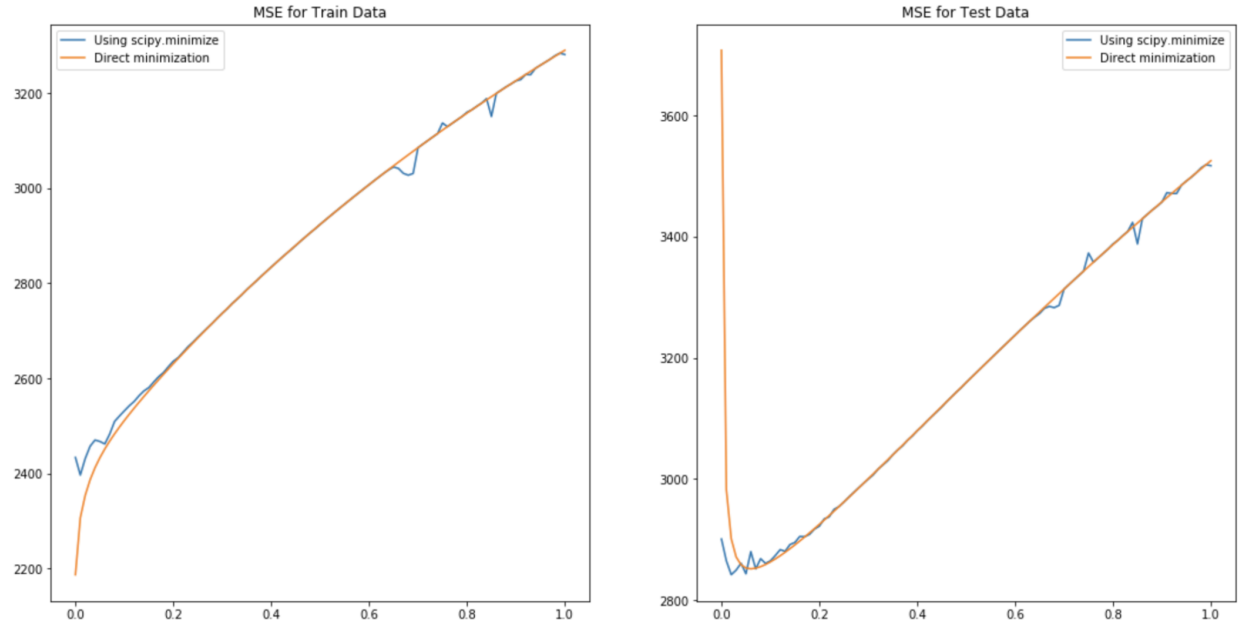
We observe that the weights of Ridge regression is low than the weights of OLE as we are regularizing in Ridge regression.

OLE Weights	Ridge Weights
[148.154876]	[150.45959807]
[1.2748521]	[4.80776899]
[-293.38352235]	[-202.90611468]
[414.7254484]	[421.7194576]
[272.08913436]	[279.45107288]
[-86639.45713932]	[-52.29708233]
[75914.46803678]	[-128.59418907]
[32341.62282634]	[-167.50057028]
[221.10121513]	[145.74068096]
[29299.55119946]	[496.30604123]
[125.23036028]	[129.94845775]
[94.41108332]	[88.30438076]
[-93.8628633]	[11.29067689]
[-33.72827999]	[1.88532531]
[3353.19771203]	[-2.58364157]
[-621.0963079]	[-66.89445481]
[791.7365324]	[-20.61939955]
[1767.76038906]	[113.39301454]
[4191.67405605]	[17.99086827]
[119.43812093]	[52.50235963]
[76.61034004]	[109.68765513]
[-15.2001293]	[-10.72779629]
[82.24245937]	[71.67974829]
[-1456.66208436]	[-69.30906366]
[827.38670282]	[-124.03437293]
[869.2909524]	[102.63981795]
[586.23449524]	[72.64220588]
[427.0267267]	[79.24754013]
[90.24676901]	[38.48319215]
[-17.88762241]	[32.98009446]
[141.69677382]	[92.09539122]
[582.8193844]	[68.97936154]
[-234.03751064]	[-24.41700914]
[-256.0714523]	[101.85387967]
[-385.1774006]	[1.39122669]
[-33.41767378]	[20.85757155]
[-10.7350066]	[-29.65490134]
[257.10718885]	[130.41115986]
[59.95545931]	[-16.75108796]
[383.72804231]	[87.51340344]
[-404.15838984]	[-45.64238362]
[-514.28643445]	[-30.92288499]
[38.36366417]	[-10.07139781]
[-44.61028891]	[31.13334896]
[-729.64353135]	[-89.33525423]
[377.4083371]	[-22.73053674]
[439.7942905]	[65.41116624]
[308.51437335]	[55.11621318]
[189.85967885]	[19.14925041]
[-109.77379702]	[-59.84315841]
[-1919.65697367]	[26.64350735]
[-1924.63377431]	[108.40501275]
[-3489.79527702]	[-137.61756968]
[11796.96874163]	[-83.04383566]
[530.67441482]	[-20.40214777]
[543.3059018]	[24.9726362]
[1821.07517995]	[-0.92451093]
[-10463.98069585]	[191.91306579]
[-516.62761094]	[34.78309393]
[2064.35917385]	[-43.90393505]
[-4199.41335621]	[23.2002376]
[-140.49570526]	[20.8504118]
[374.15709013]	[-117.853228]
[51.47574915]	[75.30611309]
[-46.44927304]	[60.36839226]

Figure 2: Weights Comparison

Problem 4 : Gradient Descent for Ridge

MSE on train and test data obtained by using the gradient descent learning by varying λ is given below :



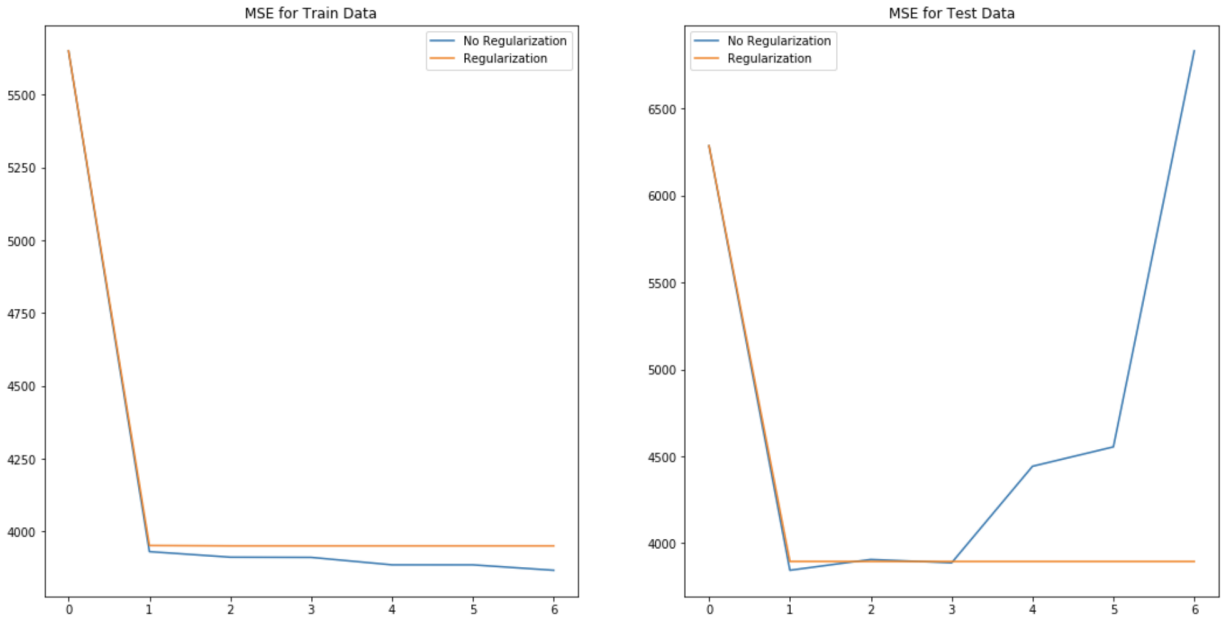
From the graph we observed that MSE is lower for gradient descent based learning.

Minimum MSE without Gradient Descent : 2851.330213443848
Minimum MSE with Gradient Descent : 2841.577430015714

Minimum λ obtained without Gradient Descent : 0.06
Minimum λ obtained with Gradient Descent : 0.02

Problem 5 : Non-linear Regression

MSE for both train and test data using regularization and not using regularization is plotted in the below graph :



We observe that without any regularization MSE decreases for train data because of overfitting and because of this overfitting we see that MSE increases for test data. MSE for both test and train data is almost constant if we use regularization.

From the plot we see that the optimal value for P is 1 in terms of test error in both setting.

Problem 6 : Interpreting Results

Our goal is to obtain minimum MSE for test data without any overfitting while learning from the train data. From our observation in problem 02 to 05 we have found that, if we use Gradient Descent for optimization with regularization paramter $\lambda = 0.02$, then we will achieve the best result. Also we have to use the data with intercept.