Introduction to Machine Learning
Project 1
Group 14
3/14/2018

Ali Alshehri
Ratnamala Korlepara

I. REPORT 1

Train both methods using the sample training data (sample train). Report the accuracy of LDA and QDA on the provided test data set (sample test). Also, plot the discriminating boundary for linear and quadratic discriminators. The code to plot the boundaries is already provided in the base code. Explain why there is a difference in the two boundaries.

The accuracy of LDA and QDA is 97% and 96% respectively. As shown in Figure 1 below, LDA and QDA boundaries are very similar as they both miss classify a few data points. The LDA boundaries are, however, slightly better as QDA boundaries go over two data points of the purple class. The LDA does slightly better here since the different classes are grouped together nicely; there isn't much overlap between them. But in QDA we are calculating covariance for each class which affected the accuracy.

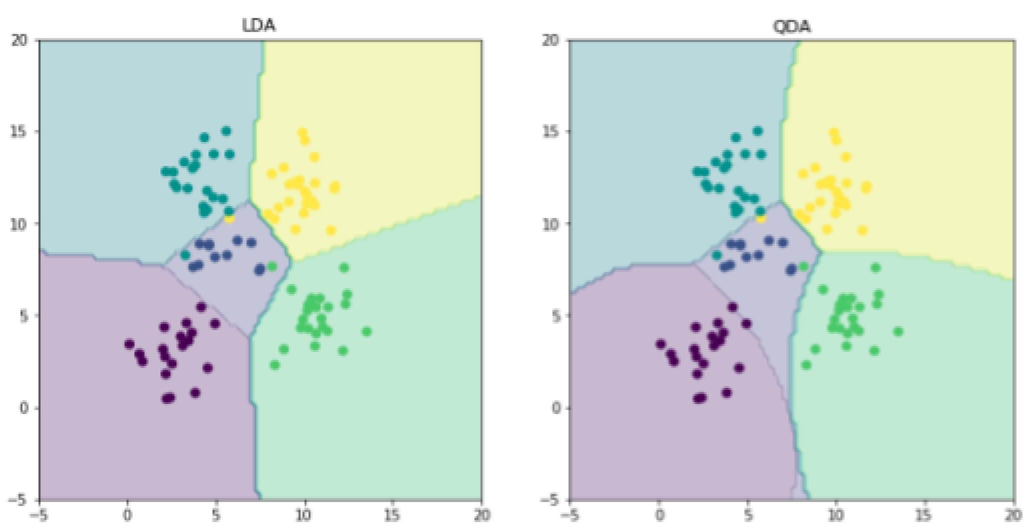

Figure 1: Class boundaries of LDA (right) and QDA (left).

II. REPORT 2

Calculate and report the MSE for training and test data for two cases: first, without using an intercept (or bias) term, and second with using an intercept. Which one is better?

MSE without intercept 106775.361555
MSE with intercept 3707.84018179

By using an intercept, we are getting lesser error value. It is better to use a bias or an intercept to get a lower mean squared error.

III. REPORT 3

Calculate and report the MSE for training and test data using ridge regression parameters using the the testOLERegression function that you implemented in Problem 2. Use data with intercept. Plot the errors on train and test data for different values of lambdas. Vary lambda from 0 (no regularization) to 1 in steps of 0.01. Compare the relative magnitudes of weights learnt using OLE (Problem 2) and weights learnt using ridge regression. Compare the two approaches in terms of errors on train and test data. What is the optimal value for lambda and why?

Table 1 below shows the relative magnitudes of weights learnt using OLE and ridge regression (λ= 0.06).

OLE_regression (problem 2)	Ridge_regression (problem 3)
-412.17	-211.24
-345.94	-196.56
578.81	543.44
58.92	391.22
-1,358,916.12	-239.51
1,194,622.63	-165.11
507,036.46	-225.58
-1,345.87	-67.45
447,713.28	283.54
477.90	322.99
-140.66	68.21
-919.34	-498.05
-395.97	-179.81
-72,569.26	-111.03
-89,509.37	-171.42

-3,237.83	-396.55
1,407.30	244.98
39,179.52	-100.72
265.08	141.02
512.84	372.23
201.16	173.08
69.91	87.13
-4,243.07	-12.54
3,446.45	-235.38
2,224.00	565.35
-177.73	-257.83
1,580.34	300.72
108.32	14.14
247.11	292.99
-23.85	-35.09
793.88	248.70
296.37	388.67
-606.95	35.37
-563.19	218.70
-548.07	-292.39
96.99	3.93
590.07	160.73
-1,343.56	-66.12
2,328.85	296.75
-159.02	22.23
-756.38	75.59
436.01	-56.87
-245.56	10.09
-8,744.83	-189.73
7,026.34	20.49
3,622.58	-6.24
578.81	165.69
3,319.48	300.28
-239.95	-172.63

158,217.35	-7.29
20,386.90	332.15
-38,610.61	-310.03
270,468.29	-452.21
5,910.68	-73.50
-24,540.71	310.90
37,194.43	152.97
-233,074.90	189.95
-5,096.87	45.64
12,994.09	-39.12
-89,589.72	-20.43
-2,499.68	-94.79
10,830.71	-354.97
-554.80	280.24

Table 1

Unlike in Problem 2 above, here we used lambda values to introduce some noise to avoid overfitting the data. As shown Figure 2 below, the model starts to underfit with higher values of lambda. Our optimal value of lambda is 0.06 because at that value we get the least mean squared error.

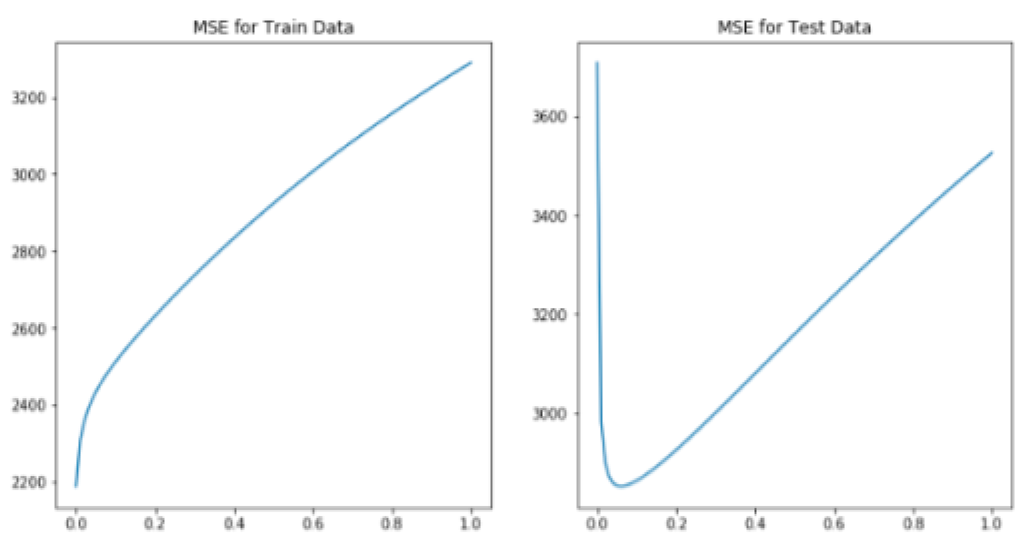

Figure 2: MSE with different values of lambda for training data (left) and test data (right).

IV. REPORT 4

Plot the errors on train and test data obtained by using the gradient descent based learning by varying the regularization parameter. Compare with the results obtained in Problem 3.

- Training data:
  MSE4: Our lowest MSE by using gradient descent to minimize loss is 2411.62573994
  Lambda for MSE4: 0.01

  MSE3: Our lowest MSE with direct minimization is 2187.16029493.
  Lambda for MSE3: 0.0

- Test data:
  MSE4: Our lowest MSE by using gradient descent to minimize loss is 2834.11841739
  Lambda for MSE4 is around: 0.06

  MSE3: Our lowest MSE with direct minimization is 2851.33021344.
  Lambda for MSE3: 0.06

For the training data, MSE is lower with direct minimization as opposed to using gradient descent (cf. 2187.16 and 2411.63 respectively). However, with respect to the test data, MSE is slightly higher when using direct minimization as opposed to using gradient descent (cf. 2851.33 and 2834.12 respectively).

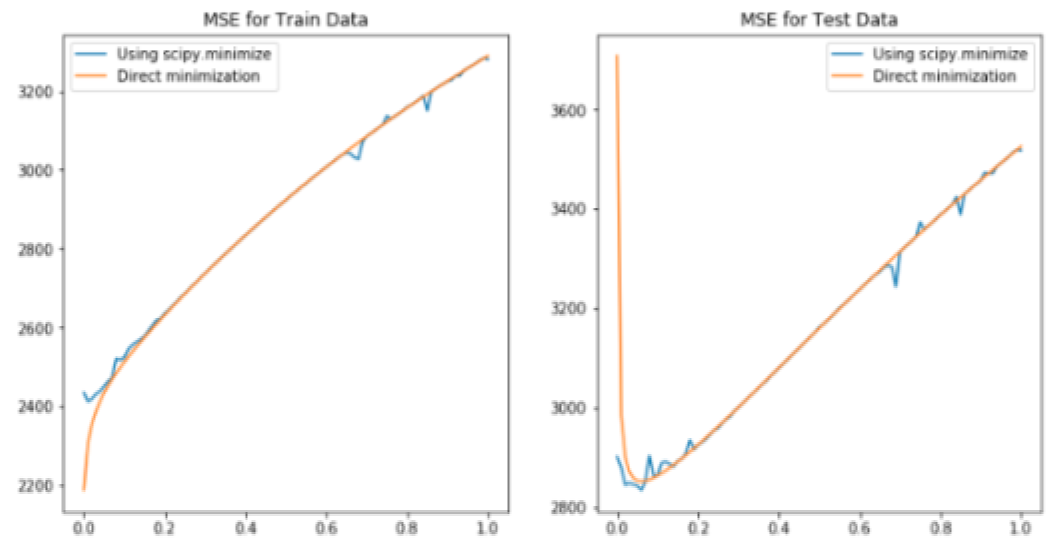

Figure 3: MSE with direct minimization and scipy.minimize for train (left) and test (right) data.

V. REPORT 5

Using the $\lambda = 0$ and the optimal value of λ found in Problem 3, train ridge regression weights using the non-linear mapping of the data. Vary p from 0 to 6. Note that p = 0 means using a horizontal line as the regression line, p = 1 is the same as linear ridge regression. Compute the errors on train and test data. Compare the results for both values of λ. What is the optimal value of p in terms of test error in each setting? Plot the curve for the optimal value of p for both values of λ and compare.

For λ=0.0:
In training data, the MSE values decreased as the p values increased from 0 to 6.
In test data, we see the opposite behavior. For p values are greater than 3, the mean squared error also increased.

For λ=0.06:
In both training and test data, the MSE values are about the same with *p* values equal or greater than 1 and very higher when $p = 0$.

Without Regularization (λ=0.0), the optimal value of p is 1 since we get the least mean squared error.

p = 0 ; MSE = 6286.40479168
p = 1 ; MSE = 3845.03473017
p = 2 ; MSE = 3907.12809911

p = 3 ; MSE = 3887.97553824
p = 4 ; MSE = 4443.32789181
p = 5 ; MSE = 4554.83037743
p = 6 ; MSE = 6833.45914872

With Regularization (λ=0.06), the mean squared error values are similar for values of p (1 to 6) but the optimal value of p seems to be 4 since we get slightly lower mean squared error.

p = 0 ; MSE = 6286.88196694
p = 1 ; MSE = 3895.85646447
p = 2 ; MSE = 3895.58405594
p = 3 ; MSE = 3895.58271592
p = 4 ; MSE = 3895.58266828
p = 5 ; MSE = 3895.5826687
p = 6 ; MSE = 3895.58266872

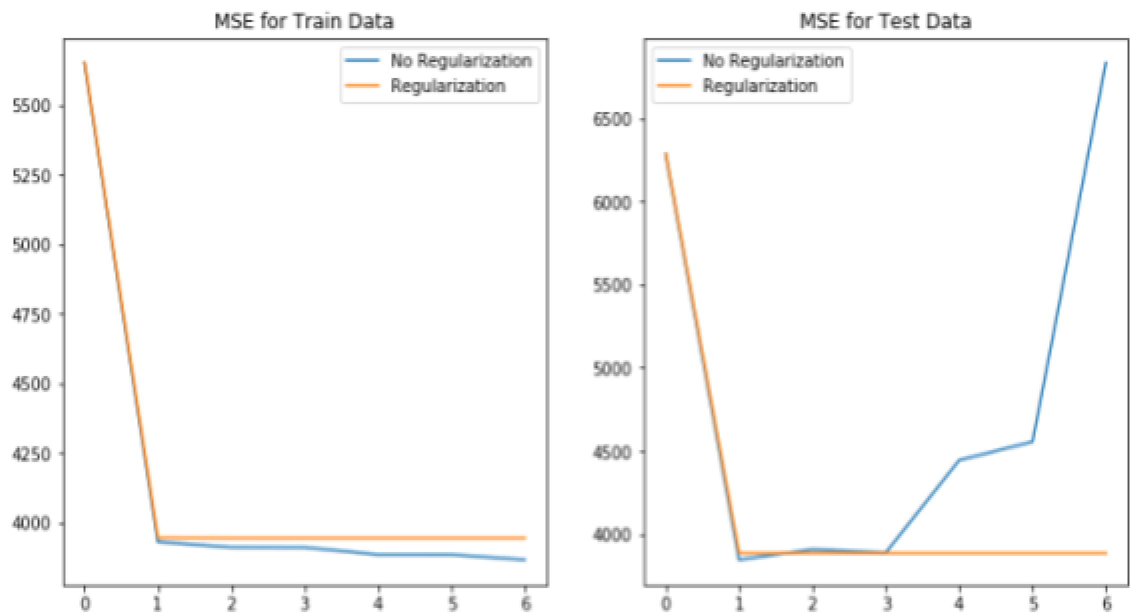

Figure 4: MSE with and without regularization (with λ = 0.06 and 0 respectively) for training (left) and test (right) data.

VI. REPORT 6

Compare the various approaches in terms of training and testing error. What metric should be used to choose the best setting?

By comparing the results of the problems above, the metrics that should be used are:

1) When calculating mean squared error, it is best to add an intercept so as to get a lower error value.
2) Using regularization such as Ridge helps to prevent overfitting of data e
3) The optimal lambda value should be the value where we get the least mean squared error.
4) We get better estimates when using Gradient Descent to minimize the loss function.
5) With respect to the test data, we get the best estimates when assuming linear relationship between the diabetes level using the input features.
6) In the light of the results as a whole, our recommendations for anyone using regression for predicting diabetes level using the input features is to use intercept, regularization as well as gradient descent for estimating the parameters.