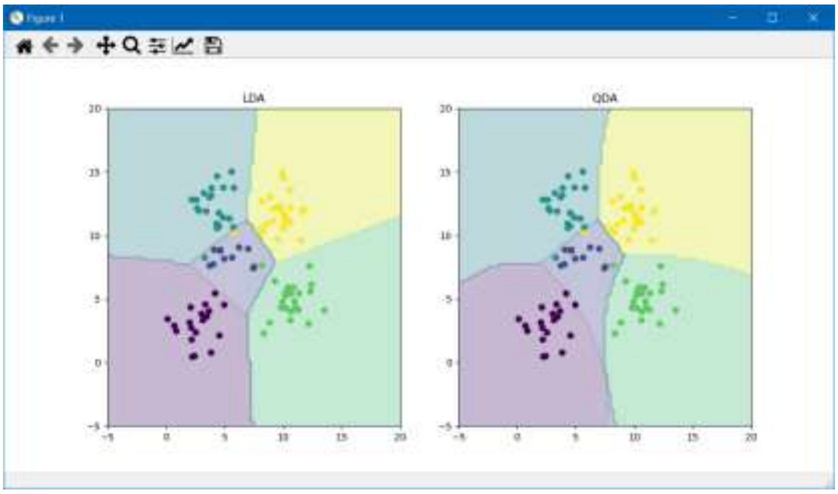
<u>Assignment 1: Report</u>

Group Members: Vinil Palavalasa, Rahul Patil, Kunal Mehta

Problem 1

a. As instructed, the sample training data in sample_train was used to train both the Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) methods. The accuracy results are:
   - LDA: 97.0%
   - QDA: 96.0%
b. The discriminating boundary for both LDA and QDA on the testing data set (sample_test):



c. The difference in the two discriminating boundaries is a consequence of the fact that LDA and QDA differ when it comes to covariance matrices for classes. LDA involves a single common covariance matrix that is common to all the classes in the dataset, whereas QDA assumes that each class has its own covariance matrix. Therefore, LDA produces boundaries that are linear in nature, as can indeed be observed in the plot above. Contrastingly, QDA produces a quadratic boundary as can once again be observed in the plot above.

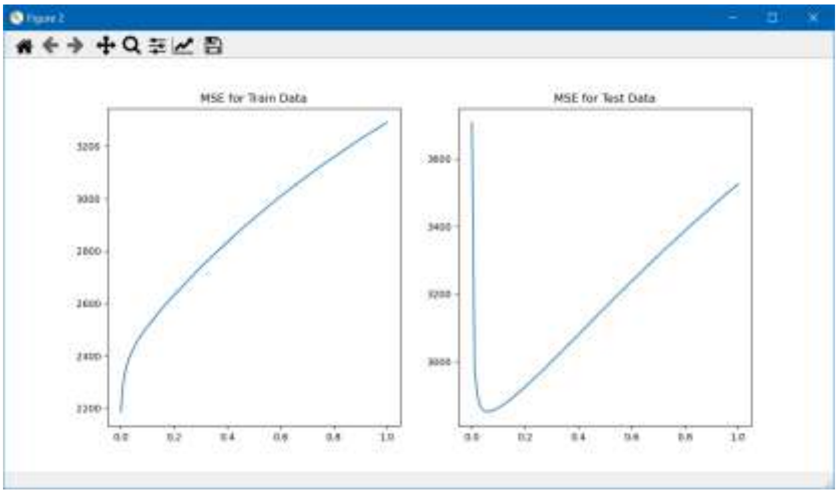
Problem 2

a. The MSE values for the data:
   - Training Data:
      i. MSE without intercept [19099.44684457]
      ii. MSE with intercept [2187.16029493]
   - Testing Data:
      i. MSE without intercept [106775.36150282]
      ii. MSE with intercept [3707.84018183]

b. The MSE for both the training and testing datasets is significantly lower when using an intercept. This is because not having an intercept forces the regression line to go through the origin which skews the line such that it is farther away from a true fit than a line with an intercept, therefore leading to a higher residuals and thereby a higher MSE value.

Problem 3

a. Plots of errors on training and testing data:



b. Optimal value of lambda = 0.06
As can be observed from the graph, the train error starts to increase from 0.01 itself with each step and does not have a downward inflection throughout the variation of the lambda values, which means that the lowest error is for a lambda value of 0.01 from which it only increases. Contrastingly, the error for the test data decreases sharply from the initial values of lambda till it reaches a minima, and then it keeps increasing. Crucially, this minima appears to be close enough to 0.01 so as to have a relatively low training data error as well. Therefore, this minima would be the ideal value of lambda. Another approach would be to compare the total of training and testing error for each value of lambda and find the value for which it is a minimum. The results we achieved are shown below. The ideal value for lambda would be 0.06 since the difference in total error between 0.06 and 0.02 is minimal and lowering testing data is much more important than reducing training error.

Based on testing data error,

- Best lambda = 0.06
- Sum of Train and Test MSE = [5302.85870409]

Based on sum of training and testing data error,

- Best lambda = 0.02
- Sum of Train and Test MSE = [5255.04493102]

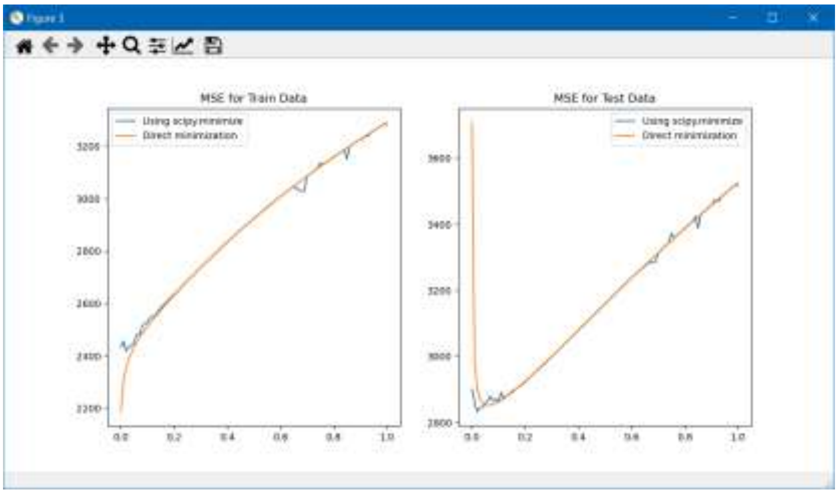
c. For weight comparison, we compared the weights of OLE Regression with the weights of regression for the ideal lambda value of Ridge Regression. The results are as shown below.

1	OLE Regression	Ridge Regression
2	[[1.48154876e+02]	[[150.45959807]
3	[1.27485218e+00]	[4.80776899]
4	[-2.93383522e+02]	[-202.90611468]
5	[4.14725449e+02]	[421.7194576]
6	[2.72089134e+02]	[279.45107288]
7	[-8.66394569e+04]	[-52.29708233]
8	[7.59144678e+04]	[-128.59418907]
9	[3.23416227e+04]	[-167.50057028]
10	[2.21101215e+02]	[145.74068096]
11	[2.92995511e+04]	[496.30604123]
12	[1.25230360e+02]	[129.94845775]
13	[9.44110834e+01]	[88.30438076]
14	[-9.38628632e+01]	[11.29067689]
15	[-3.37282800e+01]	[1.88532531]
16	[3.35319774e+03]	[-2.58364157]
17	[-6.21096288e+02]	[-66.89445481]
18	[7.91736536e+02]	[-20.61939955]
19	[1.76776039e+03]	[113.39301454]
20	[4.19167405e+03]	[17.99086827]
21	[1.19438121e+02]	[52.50235963]
22	[7.66103401e+01]	[109.68765513]
23	[-1.52001293e+01]	[-10.72779629]
24	[8.22424594e+01]	[71.67974829]
25	[-1.45666208e+03]	[-69.30906366]
26	[8.27386702e+02]	[-124.03437293]
27	[8.69290952e+02]	[102.63981795]
28	[5.86234495e+02]	[72.64220588]
29	[4.27026726e+02]	[79.24754013]
30	[9.02467690e+01]	[38.48319215]
31	[-1.78876224e+01]	[32.98009446]
32	[1.41696774e+02]	[92.09539122]
33	[5.82819383e+02]	[68.97936154]
34	[-2.34037510e+02]	[-24.41700914]
35	[-2.56071452e+02]	[101.85387967]
36	[-3.85177401e+02]	[1.39122669]
37	[-3.34176735e+01]	[20.85757155]
38	[-1.07350066e+01]	[-29.65490134]
39	[2.57107189e+02]	[130.41115986]
40	[5.99554596e+01]	[-16.75108796]
41	[3.83728042e+02]	[87.51340344]
42	[-4.04158390e+02]	[-45.64238362]
43	[-5.14286435e+02]	[-30.92288499]
44	[3.83636640e+01]	[-10.07139781]
45	[-4.46102889e+01]	[31.13334896]
46	[-7.29643532e+02]	[-89.33525423]
47	[3.77408338e+02]	[-22.73053674]
48	[4.39794291e+02]	[65.41116624]
49	[3.08514373e+02]	[55.11621318]
50	[1.89859679e+02]	[19.14925041]
51	[-1.09773797e+02]	[-59.84315841]
52	[-1.91965702e+03]	[26.64350735]
53	[-1.92463379e+03]	[108.40501275]
54	[-3.48979528e+03]	[-137.61756968]
55	[1.17969687e+04]	[-83.04383566]
56	[5.30674415e+02]	[-20.40214777]
57	[5.43305912e+02]	[24.9726362]
58	[1.82107518e+03]	[-0.92451093]
59	[-1.04639806e+04]	[191.91306579]
61	[2.06435917e+03]	[-43.90393505]
62	[-4.19941334e+03]	[23.2002376]
63	[-1.40495705e+02]	[20.8504118]
64	[3.74157090e+02]	[-117.853228]
65	[5.14757492e+01]	[75.30611309]
66	[-4.64492730e+01]]	[60.36839226]]

Ridge regression weight values are similar to that of linear regression, but the regularization of lambda = 0.06 helps in reducing the testing error and thus we see that ridge regression provides us with a lower testing error than ordinary linear regression.

Problem 4

a. The training and testing data error plots are shown below:



b. The plots for gradient descent based learning are nearly identical to the ridge regression learning plots of problem 3. Notable differences are that for very low values of lambda, gradient descent learning provides higher training errors, but lower testing errors as well. This may indicate that the optimum value of lambda for gradient descent may not be the same as that for ordinary ridge regression. Beyond that, there isn't much difference in the plots for both methods.

Problem 5

a. The errors on training and testing data is shown below:

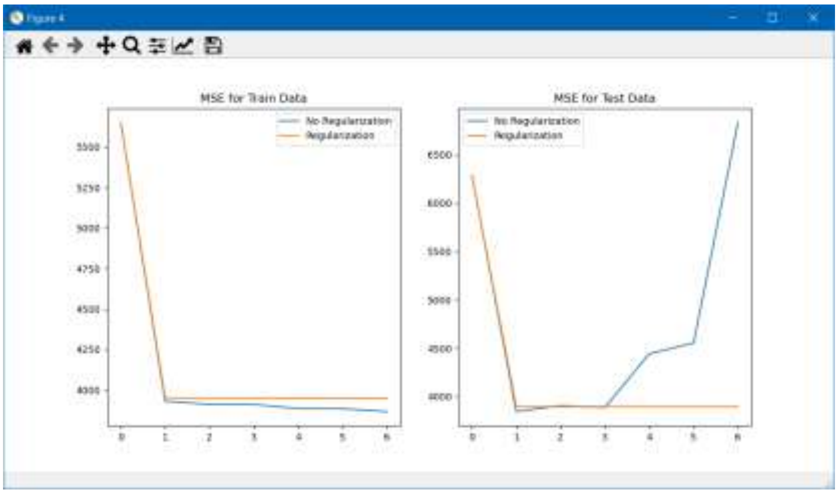
```
MSE for Train data:
 [[5650.7105389  5650.71190703]
 [3930.91540732 3951.83912356]
 [3911.8396712  3950.68731238]
 [3911.18866493 3950.68253152]
 [3885.47306811 3950.6823368 ]
 [3885.4071574  3950.68233518]
 [3866.88344945 3950.68233514]]
MSE for Test data:
 [[6286.40479168 6286.88196694]
 [3845.03473017 3895.85646447]
 [3907.12809911 3895.58405594]
 [3887.97553824 3895.58271592]
 [4443.32789181 3895.58266828]
 [4554.83037743 3895.5826687 ]
 [6833.45914872 3895.58266872]]
```

The first column represents lambda as 0 and the second represents lambda as the optimum value, which is 0.06.

For lambda = 0, the optimum value of p = 1

For lambda = 0.06, the optimum value of p = 4

The plot for optimal values of p for both values of lambda is shown below:



The plot represents our numerical findings,
For no regularization (lambda = 0), the testing error drops sharply from 0 to 1, after which it varies very less. Conversely, for lambda = 0.06, the error is more erratic, with changes occurring for each lambda value, yet supporting our numerical findings with lowest error at p = 1.

Problem 6

	Linear Regression	Ridge Regression	Gradient Descent	Non-Linear Regression
Training	[2187.16029493]	[2451.52849064]	[2417.4976429]	[3930.91540732]
Testing	[3707.84018183]	[2851.33021344]	[2832.85210624]	[3845.03473017]

a. For this table,
   - For linear regression, training and testing error values are with intercept.
   - For ridge regression, training and testing error values are with optimum lambda value of 0.06
   - For gradient descent, training and testing error values are with optimum lambda value of 0.02
   - For non-linear regression, the training and testing error values are with optimum p value of 1 with lambda as 0
b. For ridge regression, gradient descent and non-linear regression, the optimality of parameter values has been decided based on the minimization of testing error.
c. To choose the best setting, the metric must be the testing error. A lower testing error indicates higher proximity, and therefore better fit. Linear regression has a lower value of training error, but gradient descent for ridge regression learning has a lower testing error than any of the other methods, and would therefore be a better fit.
d. Therefore, based on these results, for somebody using regression for predicting diabetes level, we would recommend using gradient descent for ridge regression learning, with a lambda value of 0.02. The reason for this is that it gives the lowest testing error as compared to any of the other values.