Introduction to Machine Learning
ASSIGNMENT 1

Group members and their person number:

NAME: KARAN RAJENDRA SHAH
PERSON NUMBER: 50354506

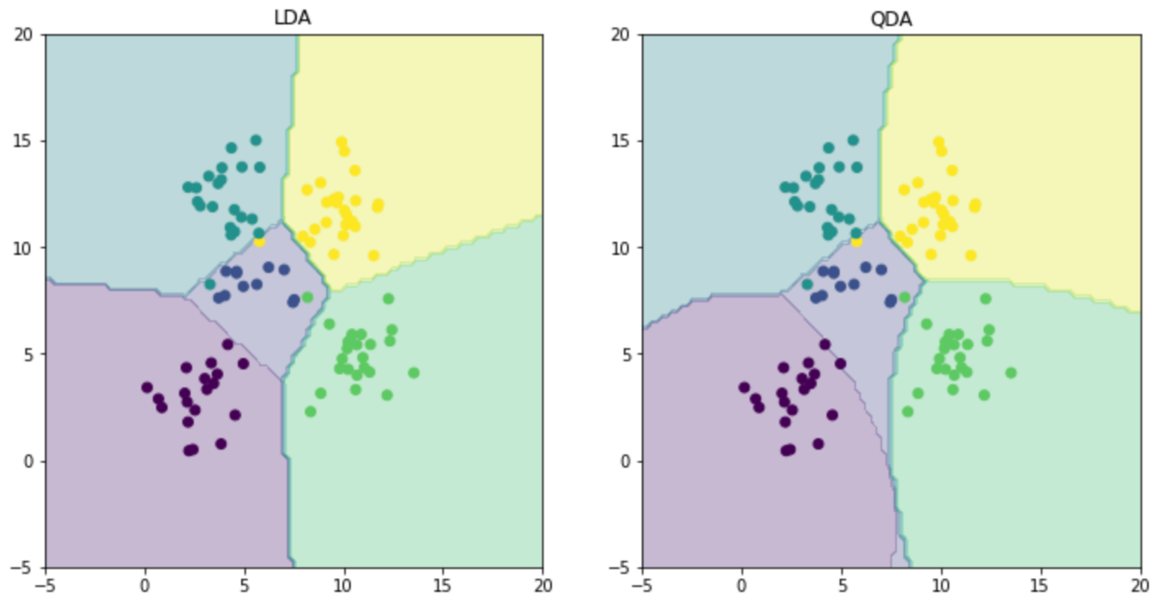
NAME: PRANAV MORESHWAR SORTE
PERSON NUMBER: 50413353

NAME: THELMA MELISSA GOMES
PERSON NUMBER: 50416894

Report 1:

The accuracy of LDA (Linear Discriminant Analysis) on test data set (Sample_test) is 0.97 or 97%
The accuracy of QDA (Quadratic Discriminant Analysis) on test data set (Sample_test) is 0.96 or 96%

```
LDA Accuracy = 97.0
QDA Accuracy = 96.0
```



LDA (Linear Discriminant Analysis) is used when a linear boundary is required between classifiers and QDA (Quadratic Discriminant Analysis) is used to find a non-linear boundary between classifiers. LDA and QDA work better when the response classes are separable and distribution of X=x for all class is normal.
The major difference between the boundaries of LDA and QDA is because in LDA we assume common covariance for all response classes i.e. covariance matrix that is common to all classes in a data set whereas in QDA we consider different covariance for each of the response class i.e. each class has its own covariance matrix.

Report 2:

```
MSE for training data without intercept[19099.44684457]
MSE for training data with intercept[2187.16029493]
MSE without intercept [106775.3615526]
MSE with intercept [3707.8401812]
```

As observed from the above screenshot, MSE for any linear regression model without the intercept is more than the MSE with intercept.

A linear regression model with intercept is better than a linear regression model without intercept.

Report 3:

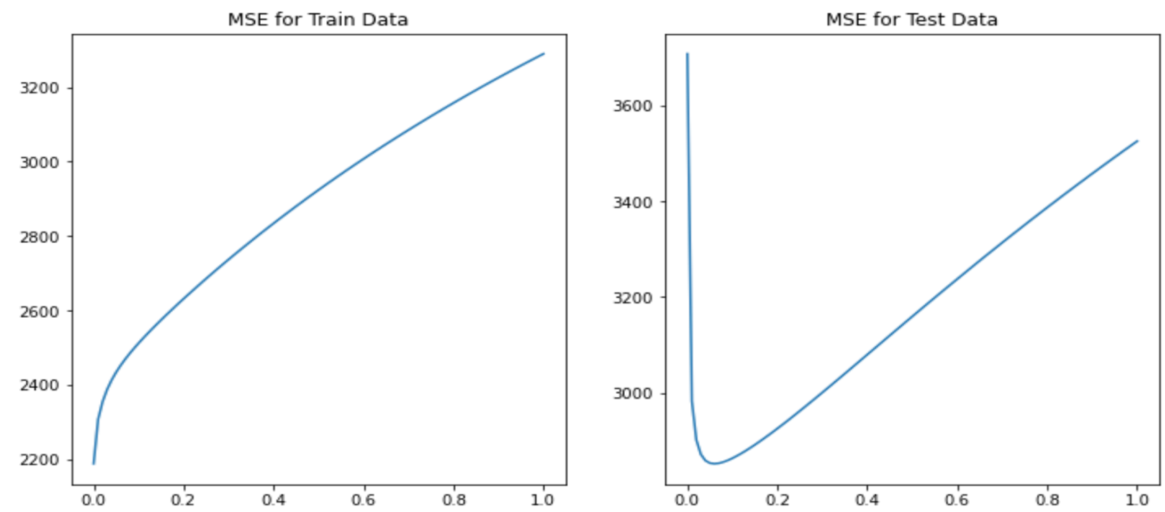
```
Magnitude of weight vector in linear regression : 124531.52652638576
Magnitudes of weight vectors in ridge regression :
[124531.52652638576, 1558.8017572659237, 1271.5989154476586, 1136.99581068829, 1056.1454854272383, 1000.703938222745
4, 959.3129608927503, 926.5390559158367, 899.4629454423923, 876.3783631906331, 856.221270408906, 838.2928408238932, 8
22.114143940337, 807.3450180449516, 793.7363974565092, 781.101065547729, 769.2950392544627, 758.205336828164, 747.74
17126489606, 737.8309331563003, 728.4127246107766, 719.4368473056383, 710.8609451744258, 702.6489395607462, 694.76981
1634782, 687.1966668674755, 679.9060072340441, 672.8771584874213, 666.0918146459379, 659.5336721124618, 653.188133074
7253, 647.0420629972078, 641.083590743397, 635.3019425924672, 629.6873034281124, 624.2306998804195, 618.923901334563
2, 613.7593355815781, 608.7300165472471, 603.829482045977, 599.0517399045222, 594.3912211127164, 589.8427389050931, 5
85.4014528735149, 581.0628373679214, 576.8226535687116, 572.6769247165789, 568.62191406889, 564.6541052198269, 560.77
01844775202, 556.9670250376764, 553.2416727315906, 549.5913331584544, 546.0133600386407, 542.5052446471769, 539.06460
6205593, 535.6891831264303, 532.3768250183584, 529.1254853715067, 525.9332148525945, 522.7981551480083, 519.718533300
3479, 516.6926564903364, 513.7189072215044, 510.7957388698568, 507.9216715649016, 505.0952883720738, 502.315231749774
45, 499.58020025705156, 496.8889454904146, 494.24026923045295, 491.63302078083956, 489.0660944840205, 486.5384273993 8
4, 484.048997131064, 481.596819793713, 479.18094810566384, 476.8004695998469, 474.45450494368873, 472.142206359987, 4
69.86275614144995, 467.61536525220794, 465.39927201016786, 463.213740844592, 461.0580611237339, 458.93154604779, 456.
83353160279216, 454.7633755714206, 452.7204565970182, 450.7041732973765, 448.71394342512565, 446.7492030717882, 444.8
094059127824, 442.8940224908532, 441.0025395355918, 439.1344593168725, 437.2892990301842, 435.4665902119816, 433.6658
7818330254, 431.88672152002226, 430.1286915482226]
```

As observed from above screenshot, the magnitude of weight vector in linear regression is 124531.5265. In the case of linear regression there is no regularization been done i.e. there is no lambda parameter used in linear regression.

In the case of ridge regression, we have the range of lambdas from 0 to 1 and it can be observed from the above screenshot that for every lambda there is a different weight vector.

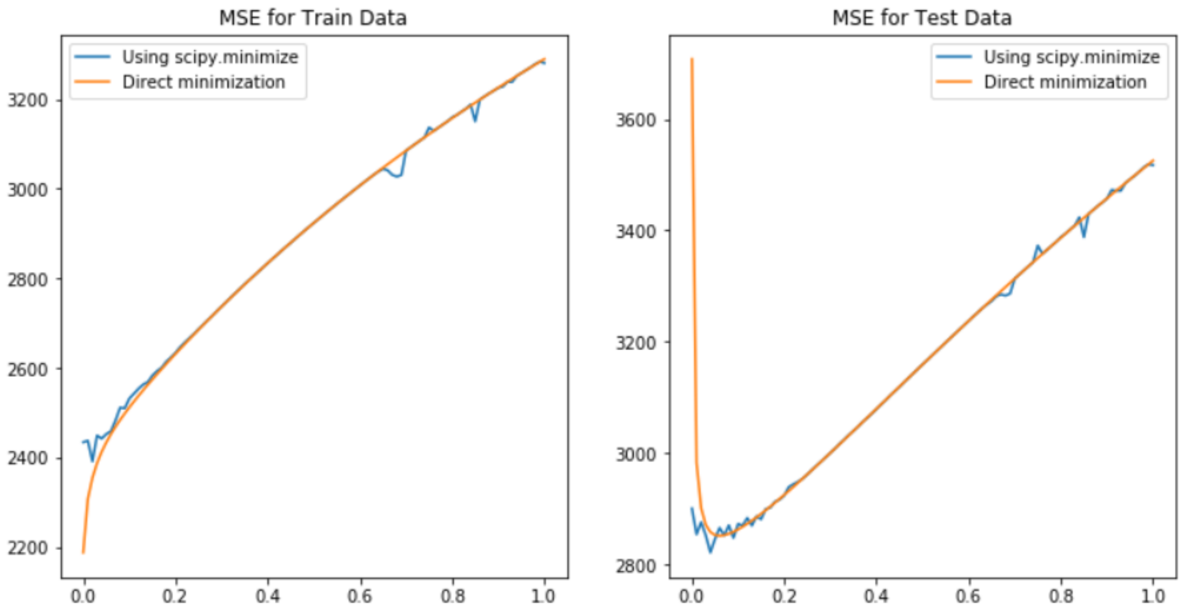
In order to find the optimal lambda, we get the lambda value in which we got the minimum MSE. In this case the optimal lambda is 0.06.

```
Optimal Lambda :- 0.06
```



The least value of MSE for training and test data using ridge regression is 2851.332021344 at lambda value of 0.06 which is less than MSE (3707.840181724566) calculated for training and test data with intercept using linear regression.

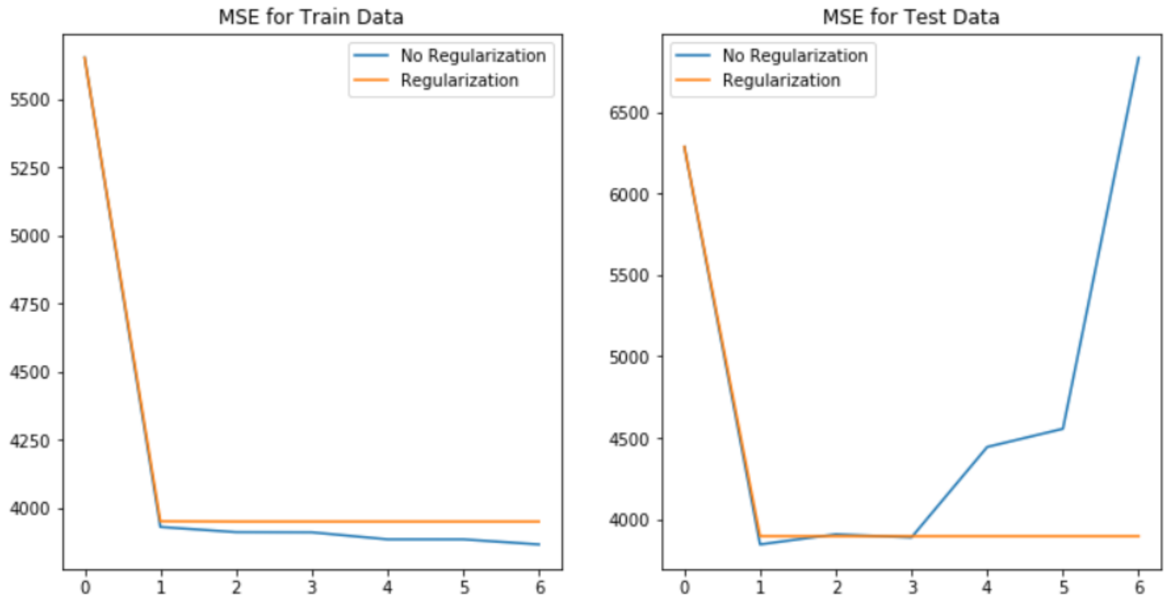
Report 4:



As seen from the above graph, we have implemented ridge regression using gradient descent. In the approach of gradient descent, we have partially differentiated the loss function and further used minimize function provided by scipy library to minimize the loss function.

In each of the subplot we have computed the MSE graph for training and testing data by using direct minimization (orange line) and scipy.minimize (blue line). We can observe the differences in both the subplots.

REPORT 5:



We have plotted the MSE graphs for lambda = 0 and lambda = 0.06 (optimal lambda obtained from Problem 3) by varying p=0 to p=6. As seen from the MSE graph of test data, we can see that optimal value of p is 1. The MSE is least when p = 1 and lambda = 0.

So we can conclude that the optimal value for p is 1 and lambda = 0.

p= 1 and lambda = 0, MSE = 3845.03473017

Report 6 –

Model	MSE
Linear Regression without intercept	106775.3615526
Linear Regression with intercept	3707.8401812
Ridge Regression	2851.33021344
Ridge Regression with gradient descent	2821.24979503
Non-Linear regression (optimal values)	3845.03473017

As per the graphs observed, Ridge regression with gradient descent gives the minimum MSE.
The MSE for ridge regression using scipy.minimize gives MSE – 2821.2497
Hence, we can conclude that Ridge regression with gradient descent will perform well as compared to other models for the diabetes dataset.