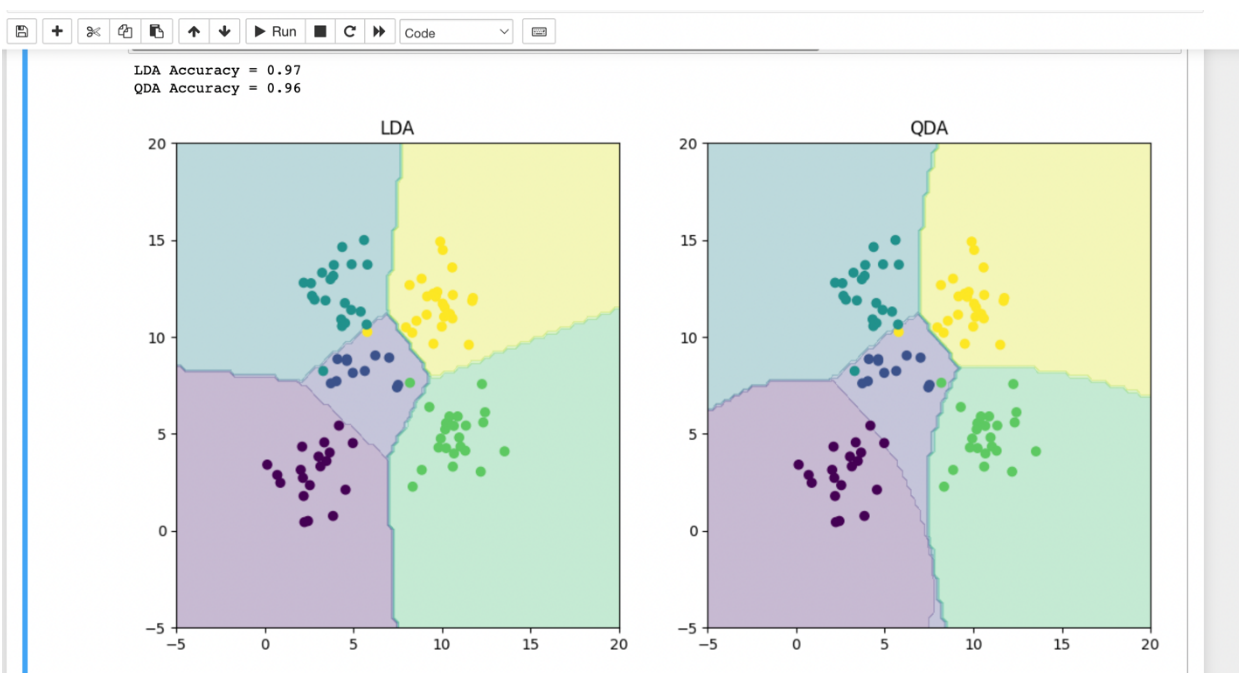
MEMBER-1: VAISHNAVI ANIL BHIDE
UBIT Name: vanilbhi
UBIT Number: 50465155

MEMBER-2: AKANKSH GATLA
UBIT Name: akankshg
UBIT Number: 50465101

PROGRAMMING ASSIGNMENT – 1

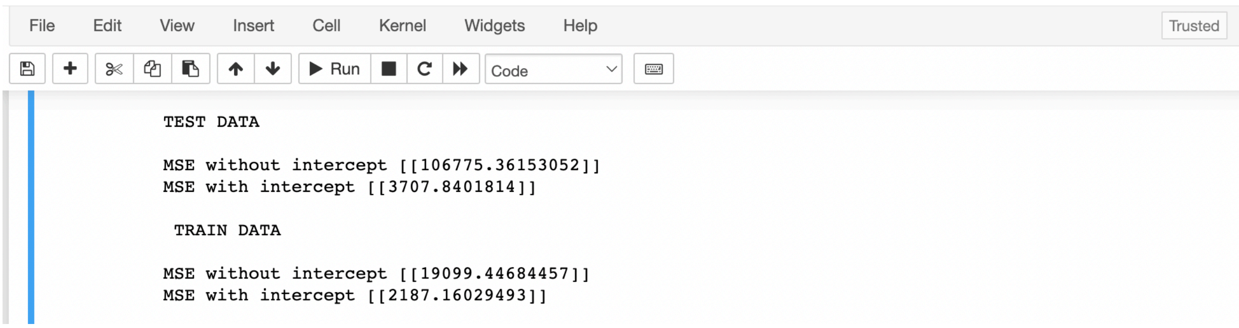
PROBLEM-1

Accuracy of LDA: 0.97
Accuracy of QDA: 0.96
From the above value LDA has higher accuracy than QDA. The decision boundary is linear for LDA. The reason being LDA considers the complete dataset while learning the covariance matrix. The decision boundary is non linear for QDA since it only considers the unique labels that are present in the dataset. The difference between the boundary line might be because of the two data points present on the boundary line in QDA.



PROBLEM -2

We can infer that the MSE change is higher for the test data when compared to the training data. Hence, using an intercept is a better option since it shows low value for test and training data with intercept




```
TEST DATA

MSE without intercept [[106775.36153052]]
MSE with intercept [[3707.8401814]]

 TRAIN DATA

MSE without intercept [[19099.44684457]]
MSE with intercept [[2187.16029493]]
```

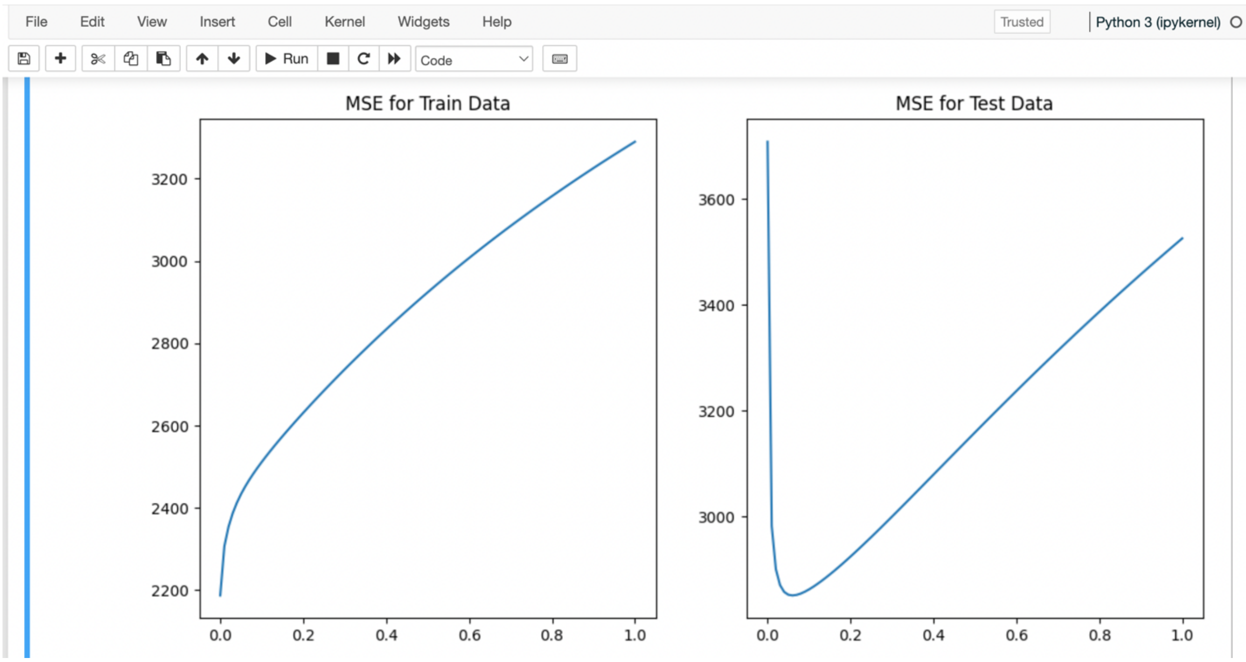
PROBLEM - 3

Th optimal value of lambda is 0.06 because MSE is lowest at this value and MSE starts increasing with lambda value beyond this point. This indicates that it's not a good fit for the dataset. As seen in the below plots,the left plot is the MSE for train data and the right plot is the MSE for test data.
We obtain the lowest MSE without regularization.

OLE and Ridge regression errors:
MSE for test data in OLE : 3707.84018177
MSE for test data in Ridge : 2851.330213
Hence, Ridge regression gives lower MSE for test data.
Weights comparison of OLE and Ridge regression: We can observe that Ridge regression has lesser weight when compared to OLE

```
ridge [[ 150.45959807]
 [   4.80776899]
 [-202.90611468]
 [ 421.7194576 ]
 [ 279.45107288]
 [ -52.29708233]
 [-128.59418907]
 [-167.50057028]
 [ 145.74068096]
 [ 496.30604123]
 [ 129.94845775]
 [  88.30438076]
 [  11.29067689]
 [   1.88532531]
 [  -2.58364157]
 [ -66.89445481]
 [ -20.61939955]
 [ 113.39301454]
 [  17.99086827]
 [  52.50235963]
 [ 109.68765513]
 [ -10.72779629]
 [  71.67974829]
 [ -69.30906366]
 [-124.03437293]
 [ 102.63981795]
 [  72.64220588]
 [  79.24754013]
 [  38.48319215]
 [  32.98009446]
 [  92.09539122]
 [  68.97936154]
 [ -24.41700914]
 [ 101.85387967]
 [   1.39122669]
 [  20.85757155]
 [ -29.65490134]
```

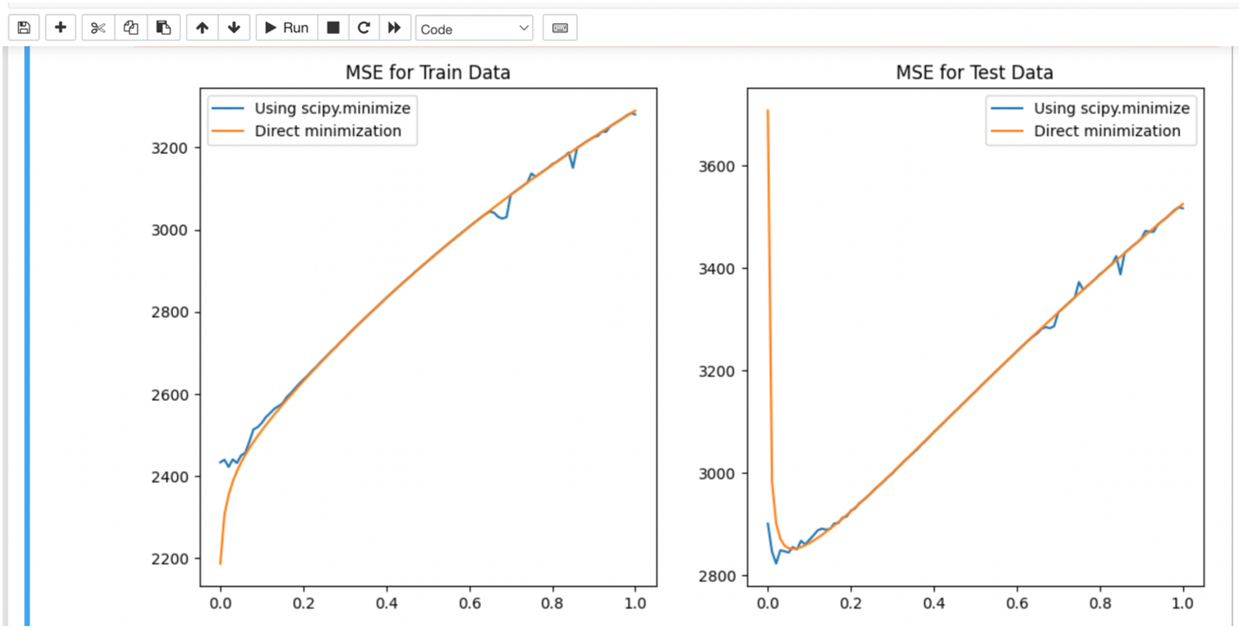
```
OLE
[[-4.12173302e+02]
 [-3.45940349e+02]
 [ 5.78814085e+02]
 [ 5.89243804e+01]
 [-1.35891612e+06]
 [ 1.19462263e+06]
 [ 5.07036456e+05]
 [-1.34586972e+03]
 [ 4.47713282e+05]
 [ 4.77903799e+02]
 [-1.40658376e+02]
 [-9.19340329e+02]
 [-3.95968904e+02]
 [-7.25692638e+04]
 [-8.95093664e+04]
 [-3.23782606e+03]
 [ 1.40730022e+03]
 [ 3.91795186e+04]
 [ 2.65084262e+02]
 [ 5.12843509e+02]
 [ 2.01158086e+02]
 [ 6.99140521e+01]
 [-4.24307012e+03]
 [ 3.44644877e+03]
 [ 2.22399630e+03]
 [-1.77734526e+02]
 [ 1.58033897e+03]
 [ 1.08315565e+02]
 [ 2.47108215e+02]
 [-2.38458154e+01]
 [ 7.93875957e+02]
 [ 2.96369204e+02]
 [-6.06952736e+02]
 [-5.63192652e+02]
 [-5.48068267e+02]
```



PROBLEM – 4

The train and test data that we have got from by using gradient descent learning by varying λ within MSE are given as:

The graph below where MSE is low for gradient descent based learning.
Minimum λ obtained without Gradient Descent is 0.06. Minimum λ obtained with Gradient Descent is 0.02. Minimum MSE without Gradient Descent is 2851.330213. Minimum MSE with Gradient Descent is 2841.577430015



<u>**PROBLEM – 5**</u>

From the above graph we can clearly determine that MSE declines for train data due to overfitting . So due to this we can clearly see that MSE in test data has an increase due to overfitting .We can infer that the value of P is one(1) for test error. In regularization, the train and test values look similar , i.e constant.

PROBLEM – 6

We need minimum MSE for test data without any overfitting while learning from the training data. From the data obtained in the previous questions, we can say that for the values that are given , Ridge Regression is better than linear Regression when the parameter regularization is around the optimal value.

The main objective is to obtain minimum value of MSE for our test data without getting overfitting. As we have seen that LDA and QDA has given the similar accuracy in classification. Here there is a covariance matrix in each class and due to this QDA is intricate. This might lead to increase in the classes and make it large. Ridge regression uses iterative inversion and transposition of matrixes is too costly and to make this considerable we have Gradient descent to reduce the overhead that is used in Ridge Regression. Hence, **Gradient descent** for optimization with regularization would give best results.