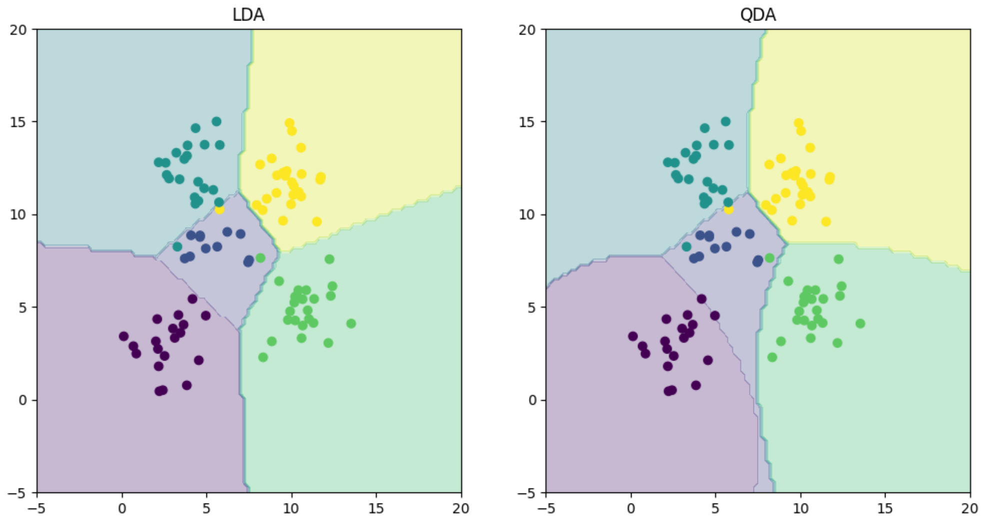
CSE 474: Introduction to Machine Learning Project 1
Group 2
Name: Yicheng Luo, Zhaoyang Pan, Yunhan Wang

Problem 1:

```
LDA Accuracy = 0.97
QDA Accuracy = 0.96
```



The reason why there is a difference in the two boundaries is, in LDA, the covariance matrix is assured to be the same for all classes. And in QDA, each class has its own covariance matrix.

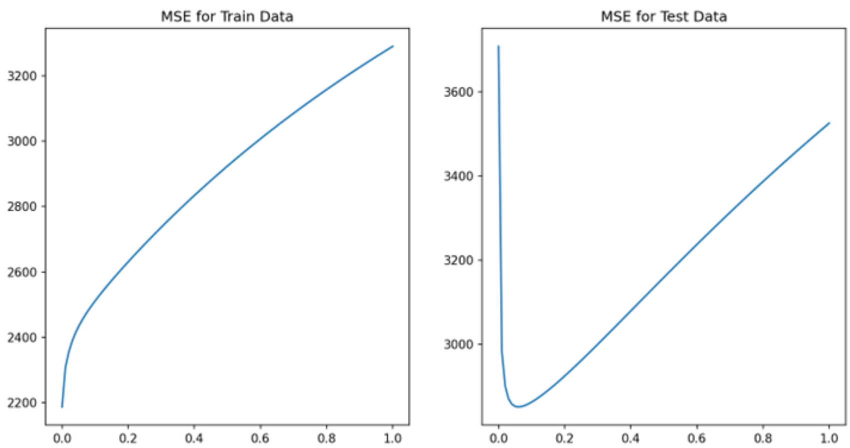
Problem 2:

```
MSE without intercept [106775.3614512]
MSE with intercept [3707.84018096]
```

As you can see in the above result, MSE is reduced from 106775.3614512 to 3703.84018096 by using an intercept. So the one with using an intercept is better than the one without using an intercept.

Problem 3:

(1) Compare the relative magnitudes of weights learnt using OLE (Problem 2) and weights learnt using ridge regression.




```
ole_w = w_i.ravel().tolist()
ridge_w = w_l.ravel().tolist()
rel_ridge_weight = sum(i**2 for i in ridge_w)
rel_ole_weight = sum(i**2 for i in ole_w)
print("The relative magnitudes of weights learnt using ridge regression: " + str(rel_ridge_weight))
print("The relative magnitudes of weights learnt using OLE regression: " + str(rel_ole_weight))
```

```
The relative magnitudes of weights learnt using ridge regression: 185010.69129298604
The relative magnitudes of weights learnt using OLE regression: 15508101011.384123
```

The relative magnitudes of weights learnt using OLE regression is greater than the relative magnitudes of weights learnt using ridge regression. Therefore, ridge regression uses smaller coefficients.

(2) Compare the two approaches in terms of errors on train and test data. What is the optimal value for λ and why?

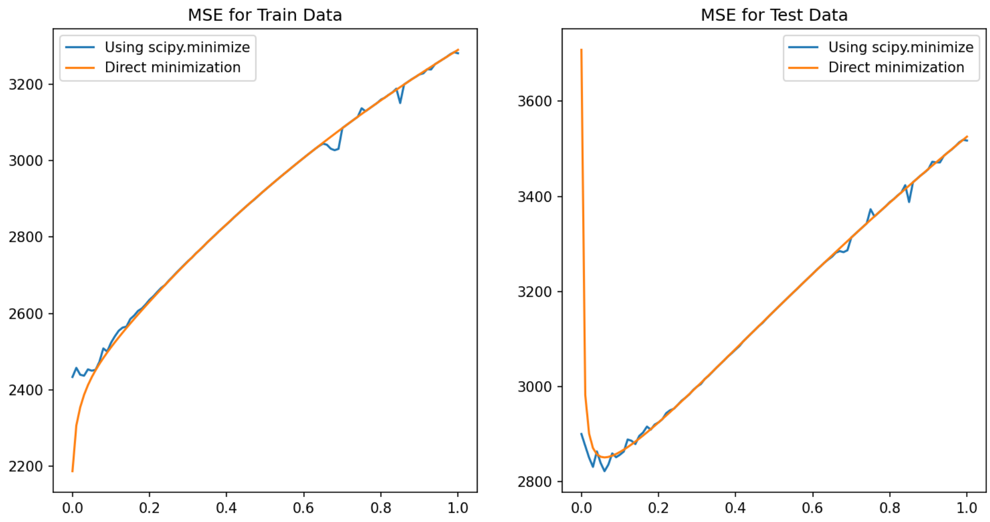
```
lambdas = lambdas.ravel().tolist()
mses3 = mses3.ravel().tolist()
min_mse = min(mses3)
min_lambda = lambdas[mses3.index(min_mse)]
print("Minimum lambda in test data:" + str(min_lambda))
```

```
The optimal lambda: 0.06
```

The optimal value for λ is 0.06. Because when the lambda value is equal to 0.06, the mse value in the test data is the smallest.

Problem 4:

(1) Compare with the results obtained in Problem 3.




```
lambdas = lambdas.ravel().tolist()
mses4 = mses4.ravel().tolist()
min_mse4 = min(mses4)
min_lambda4 = lambdas[mses4.index(min_mse4)]
print("The optimal lambda for problem4: " + str(min_lambda4))
```

```
The optimal lambda for problem4: 0.06
```

The graphs of gradient descent for ridge regression and ridge regression are almost the same. So the optimal lambda is also 0.06.

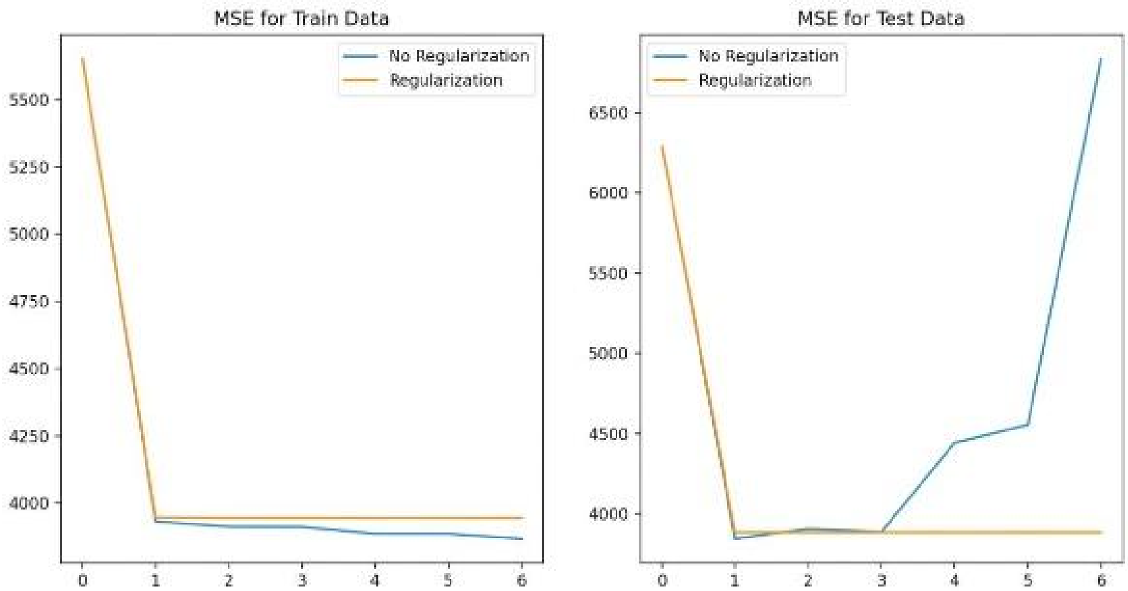
The graph of gradient descent for ridge regression will fluctuate as lambda approaches its minimum and maximum values. And because the calculations are more complex, it takes more time to generate the graph.

Problem 5:

(1) Compare the results for both values of λ, λ = 0 and the optimal value of λ found in Problem 3. What is the optimal value of p in terms of test error in each setting? For λ = 0 no regularization, the optimal value of p is 1.

For λ = 0.25 with regularization, the optimal value of p is also 1.

(2) Plot the curve for the optimal value of p for both values of λ and compare.



Problem 6:

(1) What metric should be used to choose the best setting?

According to the results above, a linear ridge regression method should be used for our dataset, with the value of parameters λ = 0.05 and p = 1.