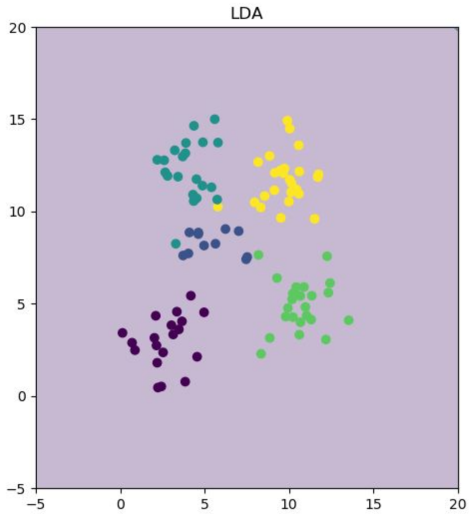
Report 2 - CSE 474 - Group 22

Hoan Duc Tran, David Olsen, Der Shen Tan

1. Problem 1.



We were unabled to get the full comparision.

2. Problem 2.

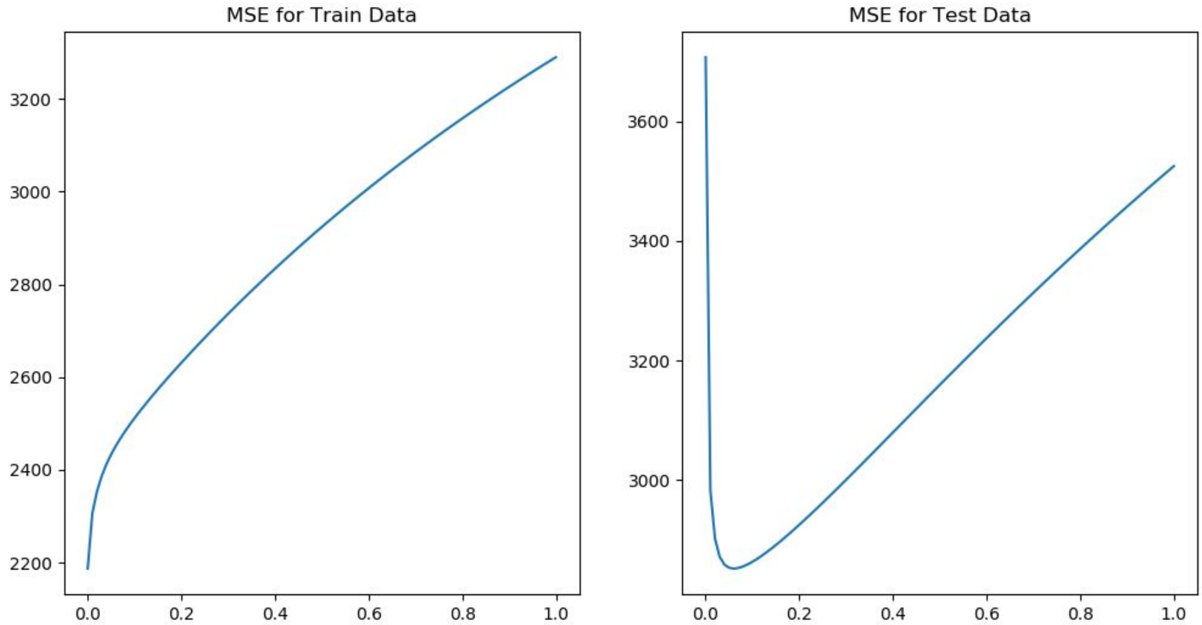
For Linear Regression using the MSE, the MSE **with intercept was much lower than without the intercept.**

MSE without intercept 106775.36150046131
MSE with intercept 3707.840181703082

Therefore, it is concluded that **using intercept would yield a much better result.**

3. Problem 3.

The MSE error for different values of lambda for test and training data can be found below.



As you can see, with lambd= 0, the MSE result is at roughly 3700, which is exactly similar to linear regression.

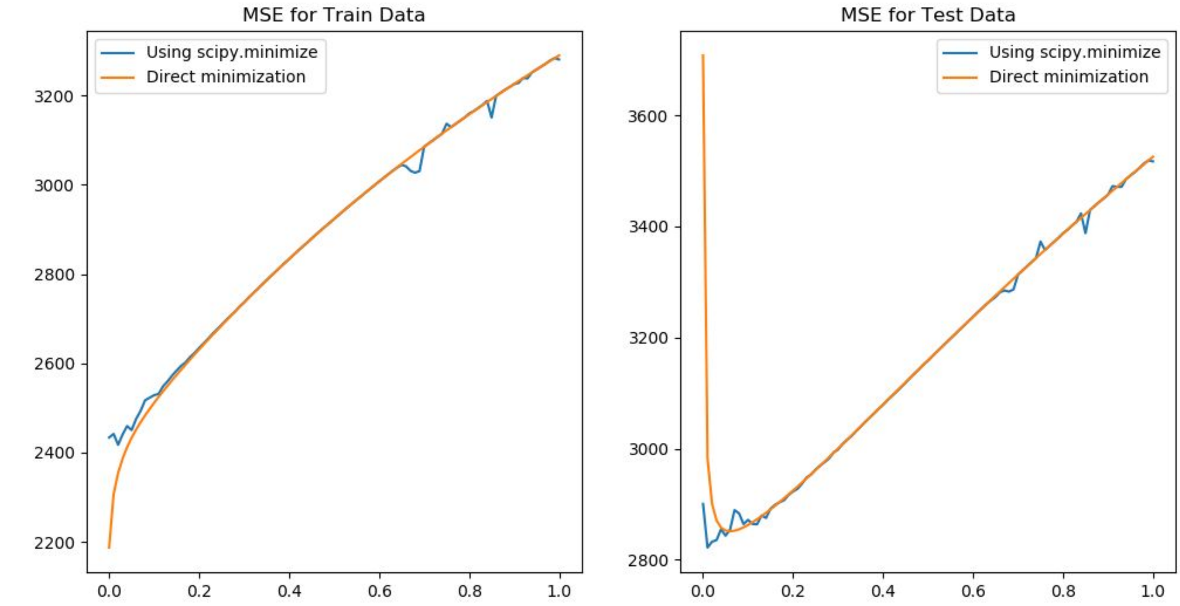
Magnitude wise, with **lamd = 0**, the sum of weights for both **linear regression and ridge regression are similar** to the 5th significant figures. At 57382.495485900254 for linear regression, and 57382.495486553285 for ridge regression. With higher lambdaval, in particular **lambd = 0.06** (this is the best value as explained below), **the sum of weights is lower at 2082.933873872311, which is almost a third of linear regression. Magnitude wise,** we could say that **Ridge regression allows for smaller weights**, with higher accuracy.

With ridge regression, we can introduce a lambd value to regularize (avoid overfitting). Thanks to lambd, the MSE for test data improve drastically, but with too much regularization, the accuracy drops down. It is expected that with higher lambd, the MSE for train data would increase.

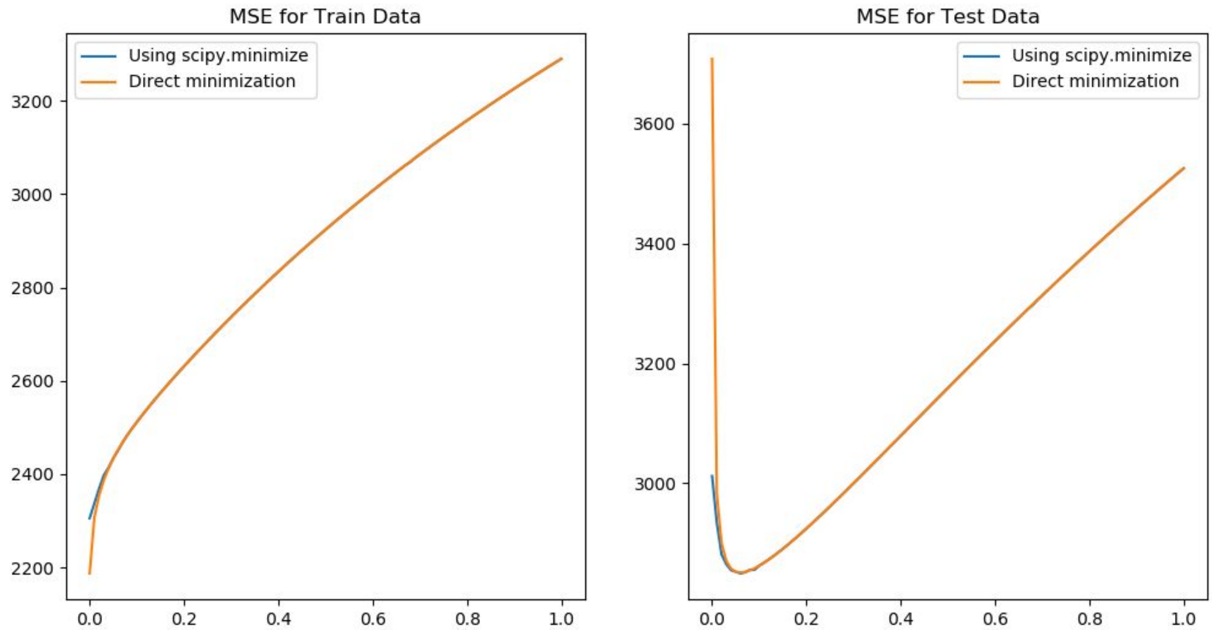
The **best MSE we have is 2851.3302134438477** at **lambd = 0.06**.

4. Problem 4.

The result is as below.



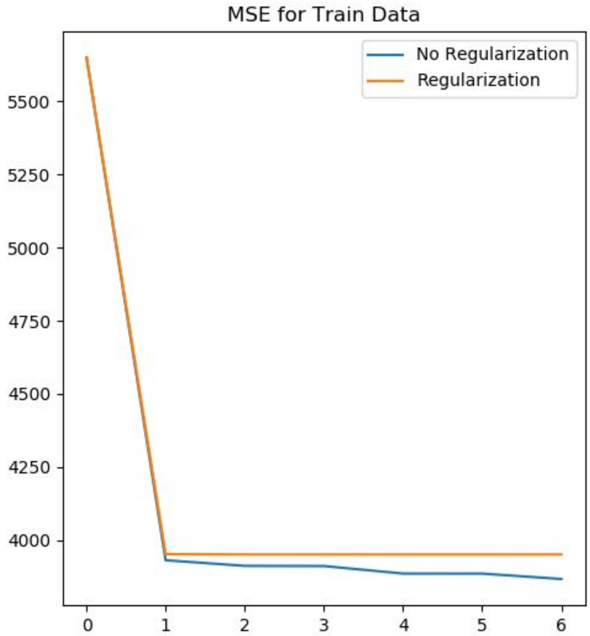
The **scipy.minimize function** kinda **overshadow the direct minimization** that we have, but not completely. We believe that this might be a limitation of max_iter, as it does not allow for better convergence on gradient descent. We changed the max_iter to 50, result as below.

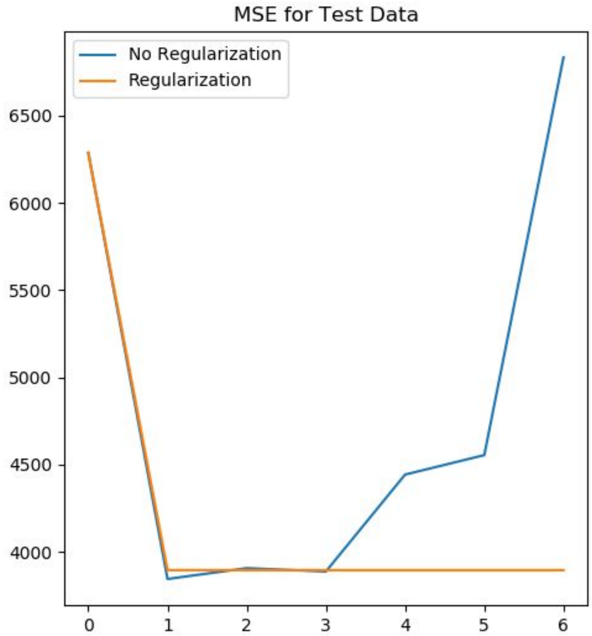


The two function almost completely overlap. This suggest that it is very likely that both implementation were correct.

5. Problem 5.

The result is demonstrated below.





For both data, **the optimal (lowest MSE) seems to be similar**. Another similarity is at **p=0, both have really high MSE**, suggest that the accuracy is really low. For higher value of p, the train data give consistently low result. Nevertheless, for test data, without regularization, the MSE increasingly worsen as the value of p increase. Regularization seems to drastically lower the error rate for test data at higher polynomial degree. Overall speaking, the optimal value if **p = 3, for both the train and test data, with or without regularization.**

6. Problem 6.

Overall speaking, the **optimal regularization value is 0.06**. And an **intercept term is recommended.**

Since the diabete data has multiple input dimension (each input has multiple features), therefore using Ridge regression clearly has an advantage. As demonstrated in problem 2 and 3, **the optimal result for Ridge regression is much better than Linear regression**.

One could use polynomial regression to expand the data. We do not have the full result to compare (it could be possible that we choose to expand certain, or all dimension of the input for better result). But if one is to expand only the third feature of each input, the result for **polynomial regression would not be as good as using Ridge**, or even Linear regression.