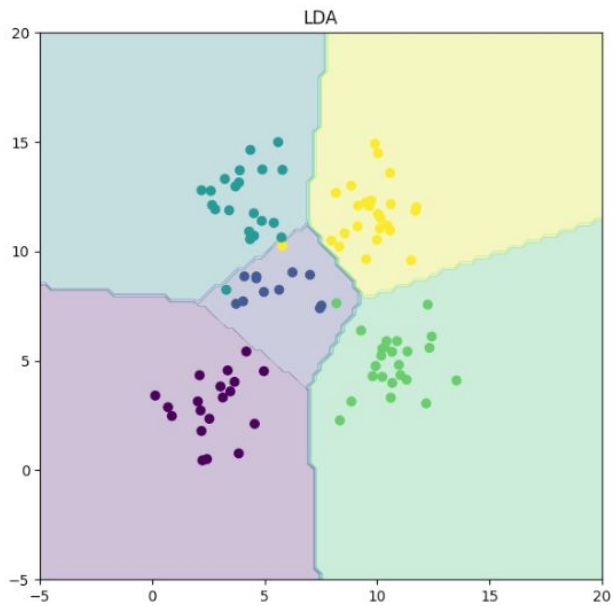
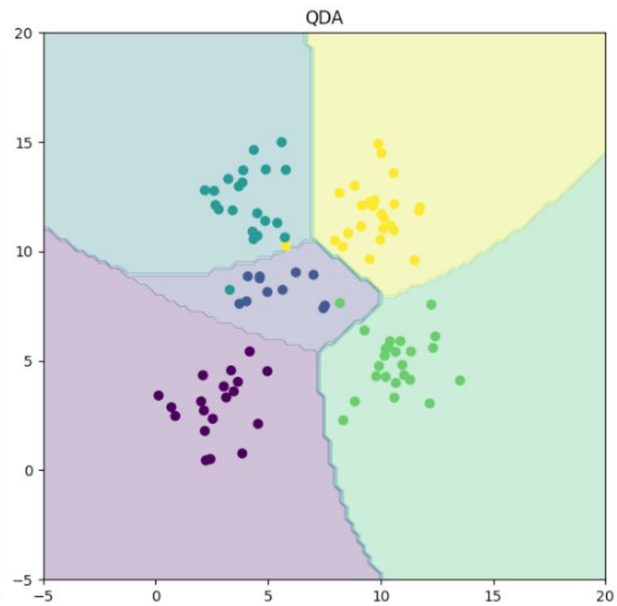
CSE474 - Classification and Regression HW1

By: Kyle Shuttleworth, Calvin Kyi, Yuanyi Yang

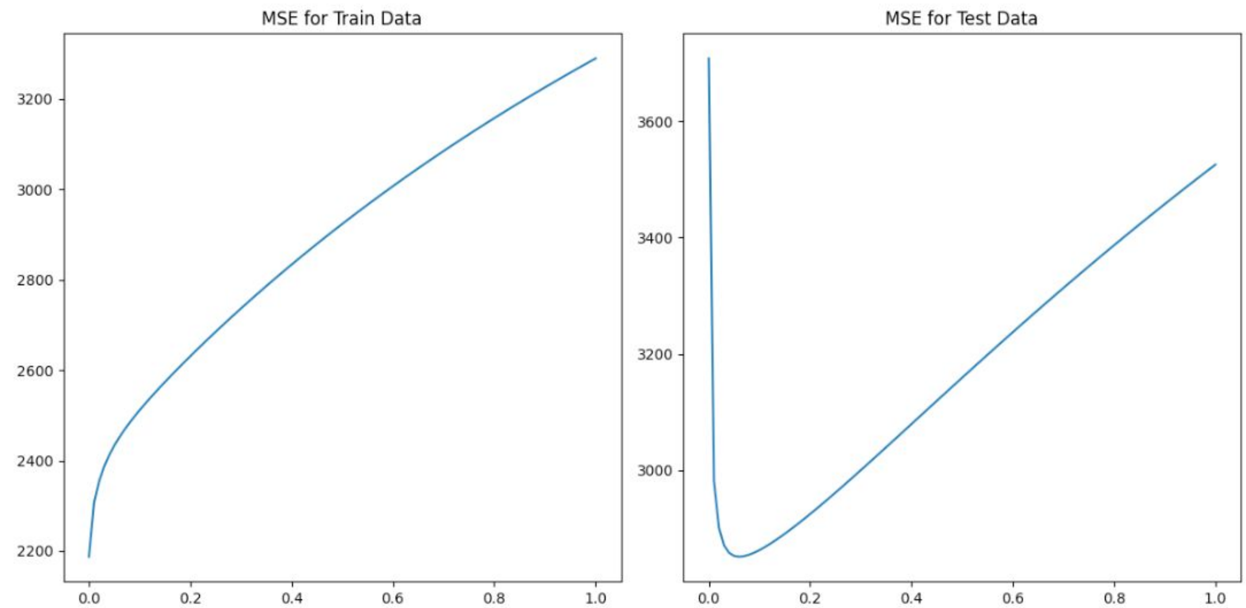
Report - 1: The accuracy of LDA on the provided test data set is 97% and the accuracy of QDA on the provided test data set is 96%. There is a difference between the two boundaries because LDA outputs a linear discriminant boundary based on the assumption that observations vary consistently across classes. However QDA doesn't make this assumption so when the variability between the classes differ it is able to capture the differences in covariances between the classes and provide a more accurate discriminant boundary. QDA is able to provide this because it assumes that each class has its own covariance matrix.





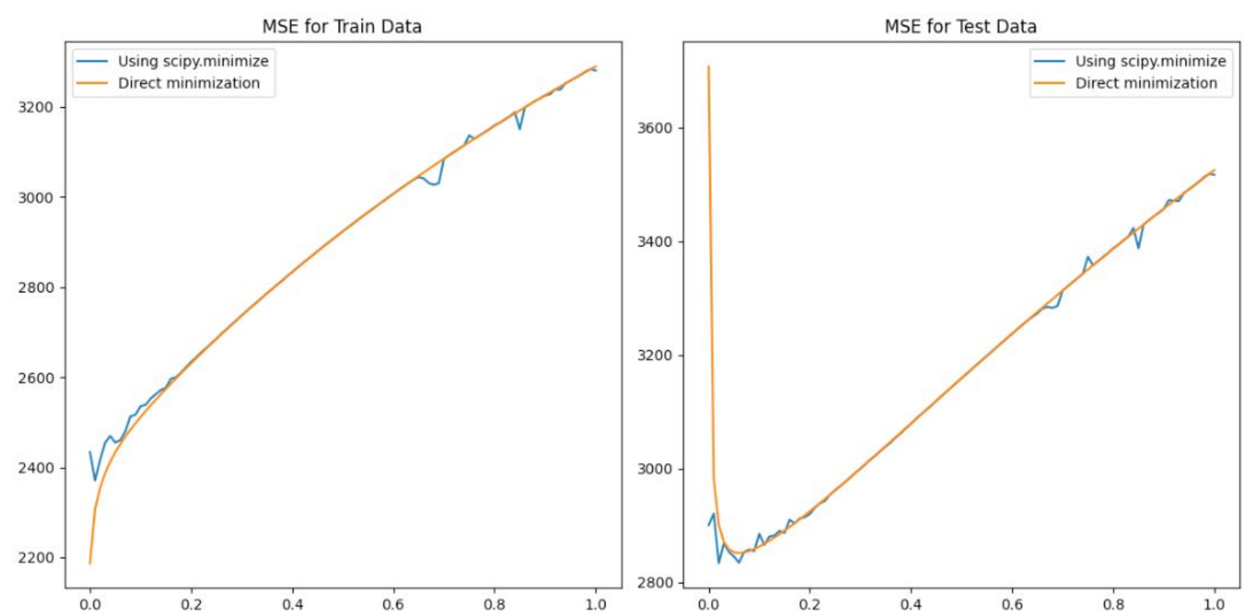
Report - 2: The mse for the test data with intercept is 3707.840181409167 and without intercept is 106775.36156629925. The mse for the train data with intercept is 2187.160294930382 and 19099.44684457022 without intercept. The mse with intercept is better because we get smaller mse that more closely maps to the points.

Report - 3: This is the graph obtained by plotting lambdas onto x-axis and mse's onto y-axis for training and test data:

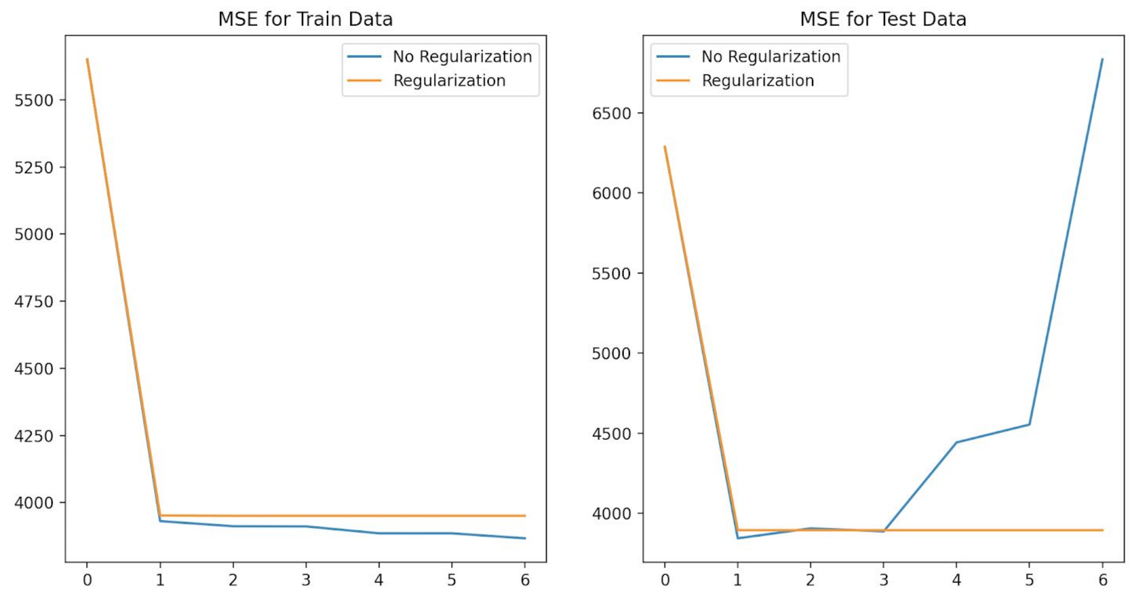


As we can see through the graph, the optimal value of lambda, which where we can get the smallest value of mse, are different for training and test data. For training data, the optimal value of lambda is 0 because mse goes up as our lambda does. On the other hand, mse approaches a very high value as the lambda goes 0, so the optimal value of lambda is approximately 0.03.

Report - 4: The graph we get from problem 4, which is using the function scipy.minimum maps more specifically than the graph we get from problem 3. While the results are more precise, it is hard to tell which one is better in general. As you can see through the graph, problem 3 is more suitable than scipy minimum in this case since it's less likely to get affected by extreme data, in other words, a steadier line that helps us more on finding the optimal value of lambda.



Report - 5: The optimal value of p when lambda = 0 is 6 for the train data and 1 for the test data. The optimal value of p when lamda = .06 (optimal value of λ found in Problem 3) is when p is greater than or equal to 1 for both the test data and the train data. Lambda = 0 is better for the train data because it is less than the mse for p compared to when lambda = .06 when p is greater than 1. Lambda = .06 is better for the test data. Its mse remains the same for p greater than or equal to 1. This is compared to lambda = 0 where mse drastically increases for p greater than or equal to 3 due to no regularization. There are a couple points where p is small where lambda = 0 has a lower mse than lambda = .06 but for the vast amount of possible p's for the test data lambda = .06 is better.



Report - 6: We believe ridge regression from problem 3 should be the best metric. This is because It has a very similar or lower mse compared to other metrics and uses a regularization process so it is less sensitive to outliers in the data. We didn't choose linear regression from problem 2 because although it has a similar or lower mse compared to ridge regression it doesn't have a regularization process so it can be much more affected by outliers which could make our regression less accurate. We didn't choose gradient descent of ridge regression from problem 4 because although it has around the same mse it is less accurate then ridge regression due to the tradeoff it makes by going for less computational cost vs higher accuracy. We didn't choose non linear regression from problem 5 because it has a higher mse then regression. We believe in this case the data is better mapped by a linear regression algorithm