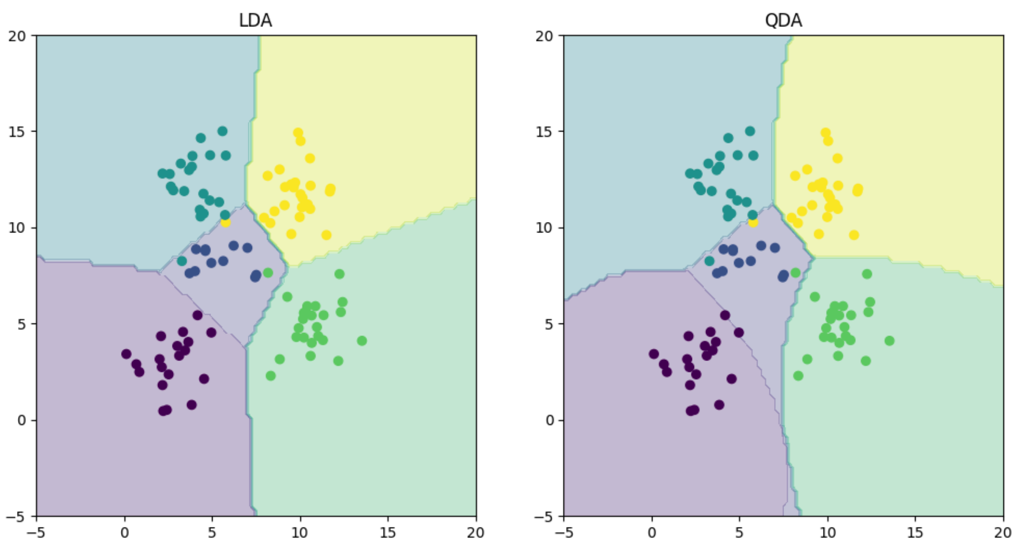
CSE 4/574
Gaussian Discriminant Analysis and Linear Regression

Rutuja Badve 50604168 (rutujara@buffalo.edu)
Ramachandran Kulothangan 50611273 (rkulothu@buffalo.edu)
Rudraksh Agarwal 50604938 (rudraksh@buffalo.edu)

Problem 1

LDA Accuracy = 0.97
QDA Accuracy = 0.96



Why is there a difference in the boundaries?

LDA and QDA are both dimensionality reduction techniques.

LDA finds a linear combination of features that separates classes within the dataset. The decision boundary between 2 classes is always linear, so in the above LDA graph, we observe straight lines as boundaries between all classes.

Unlike LDA, QDA assumes that each class can have its own individual covariance matrix. QDA has a quadratic term in its discriminant, which results in curved shapes of the boundaries between different classes.

Problem 2

$w = (X^T X)^{-1} X^T y$

Train data

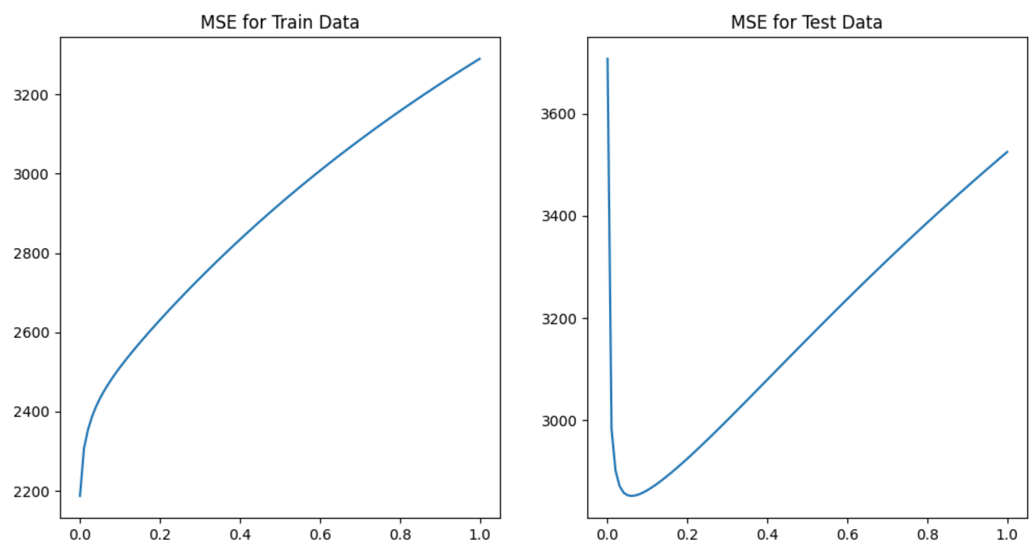
MSE without intercept: 19099.446844570746
MSE with intercept: 2187.1602949303892

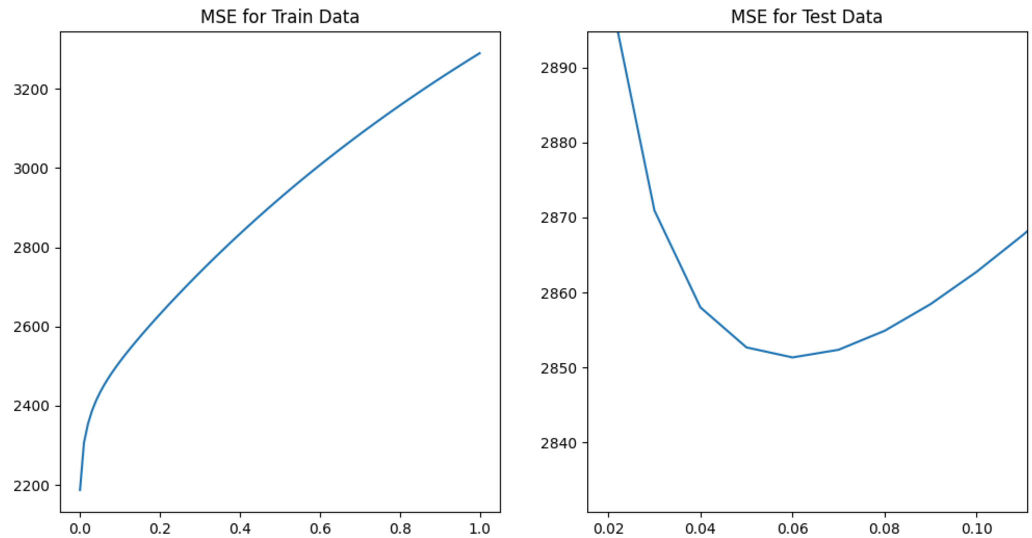
Test data
MSE without intercept: 106775.36153972965
MSE with intercept: 3707.8401811277313

In both the datasets (train and test), MSE with intercept is significantly less than MSE without intercept, and thus MSE with intercept is better. Thus, intercept should be used for linear regression for this case.

Problem 3

$w \ = \ (\lambda I \ + X^T X)^{-1} X^T y$





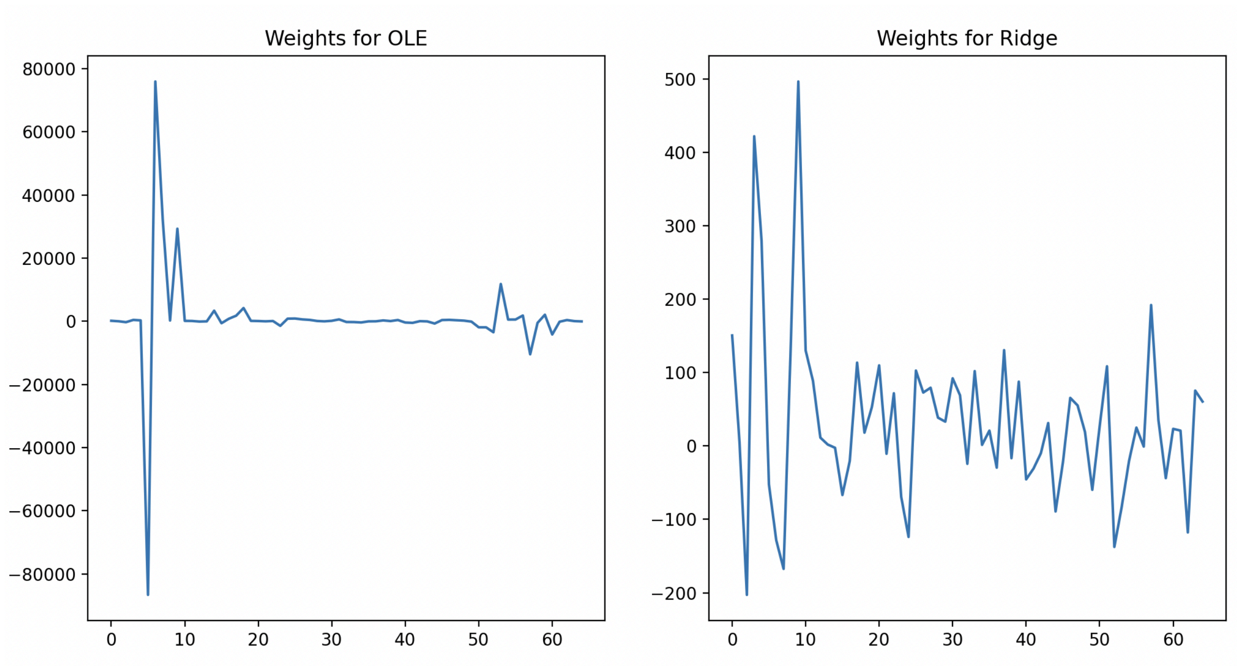
Optimum Lambda
The optimum value of lambda can be found by noting the value of lambda for which the MSE is minimum.

lambda_opt = 0.06 (at this value lambda, MSE for test data is lowest (~2851)

Comparing Errors for training and test data
We can see that the error for training data is lower with OLE compared to Ridge regression, but for test data, we find Ridge regression to give a much lower error as compared to OLE.

MSE for train data OLE = 2187.16
MSE for train data Ridge (Lambda = 0.06) = ~3000

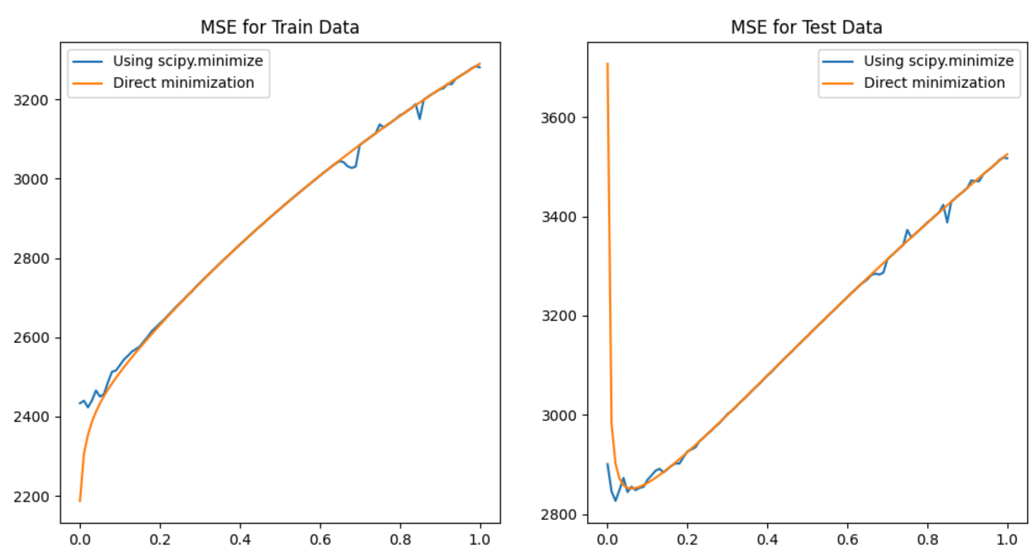
MSE for test data OLE = 3707.84
MSE for test data Ridge (Lambda = 0.06) = ~2851

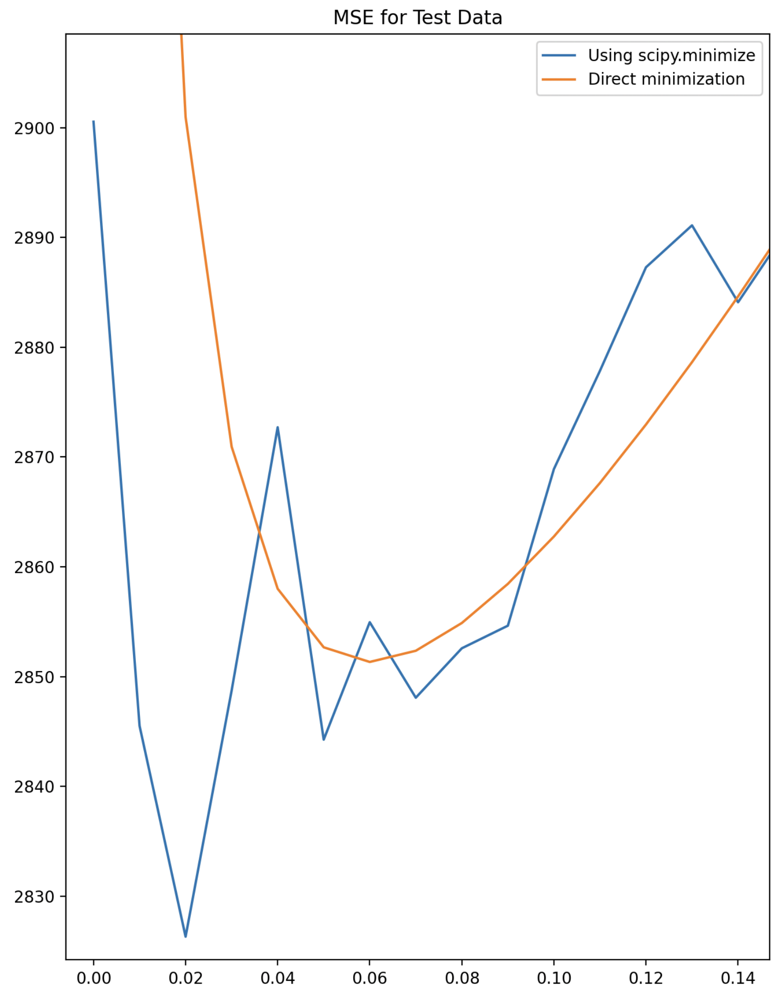


Comparison for weights

We can see that in the case of OLE, the values of weights were huge and varied (-80000 to 80000), but in the case of ridge regression, the values of weights were comparatively relatively small (-200 to 500). This means in ridge regression, weights are heavily penalized.

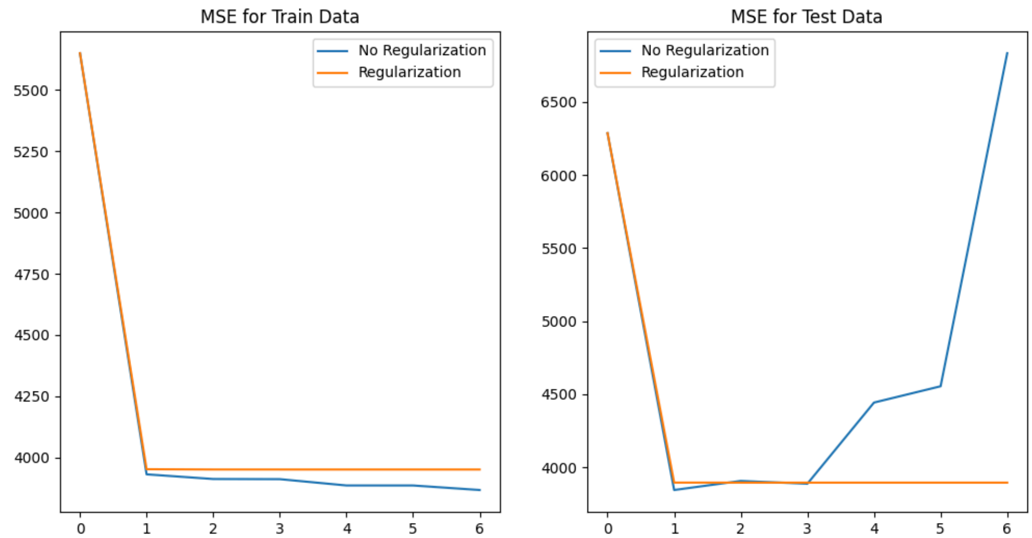
Problem 4





The best MSE for test data for regularization using gradient descent is slightly lower than the direct minimization. This tells that there exists a better weight with a different regularization parameter λ, which makes the MSE go down by a bit. However, the difference in MSEs is insignificant compared to MSE absolute values.

Problem 5



We can see that for train data when $\lambda = 0$, the MSE continues to decrease for increasing values of p, we can see overfitting in our model. With $\lambda = 0.06$, MSE is consistent for p>=1 and is constant and avoids overfitting.

But for test data, when $\lambda = 0$, the MSE is minimum at p = 1 and then after p = 3, it increases significantly for every increase in p due to overfitting, with $\lambda = 0.06$, MSE remaining constant for p >= 1 and staying at a minimum.

For $\lambda = 0$, the optimum value of p = 1,
For optimum, $\lambda = 0.06$, the optimum value of p = 1

Problem 6
Findings:

1. In ordinary linear regression, we found using intercepts to be more suitable for our dataset as MSE was significantly less for testing and training data.
2. In our case, non-linear regression is not preferable, as we observed that we get the best solution for p=1(i.e., linear) as compared to any polynomial regression. The MSE for training data reduced on increasing p, but the MSE for testing data increased a lot on increasing p.
3. Ridge regression gave the best solutions for our dataset, using a regularization parameter of 0.06. It also reduced the weights to be in a much smaller range[-200,500] as compared to OLE[-80000,80000].
4. Ridge regression using gradient descent gave the best results with a regularization parameter of 0.02, where the MSE for test data was lowest(~2825).

Notes:

1. We should do cross-validation and use different subsets of training data to find the best weights.

Recommendation:

Based on the above findings, our recommendation is to use **Ridge Regression with gradient descent** and a regularization parameter of 0.02. This is using MSE for test data as our deciding metric.