CSE 574-Introduction to Machine Learning
Programming Assignment – 1

Group 24

Aboli Rawanhate(aboliraw@buffalo.edu)

Manasa Gajulavarthy(mgajulav@buffalo.edu)

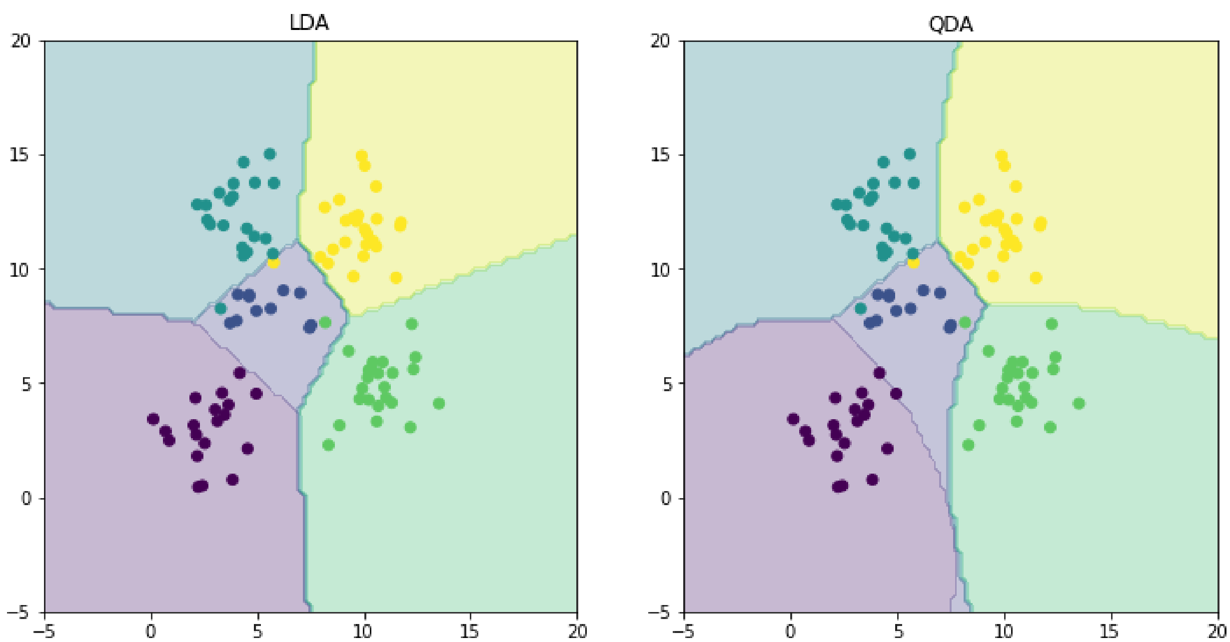
Vamsi Krishna Kokkiligadda(vkokkili@buffalo.edu)

Problem 1: Experiment with Gaussian Discriminators

Accuracy of LDA and QDA on the provided test data set (sample test):

LDA Accuracy = 0.97
QDA Accuracy = 0.96

Discriminating boundary for linear and quadratic discriminators:



Observations:

- From the above plots, it is evident that both QDA and LDA discriminate data into 5 classes.
- LDA has linear boundaries as it doesn't have the quadratic term. Also, instead of calculating the covariance matrix for each class, we use the entire dataset to calculate covariance matrix for LDA
- In QDA, the boundaries are quadratic. Also, covariance matrix is calculated for each class.

Problem 2: Experiment with Linear Regression

MSE for training and test data for two cases:

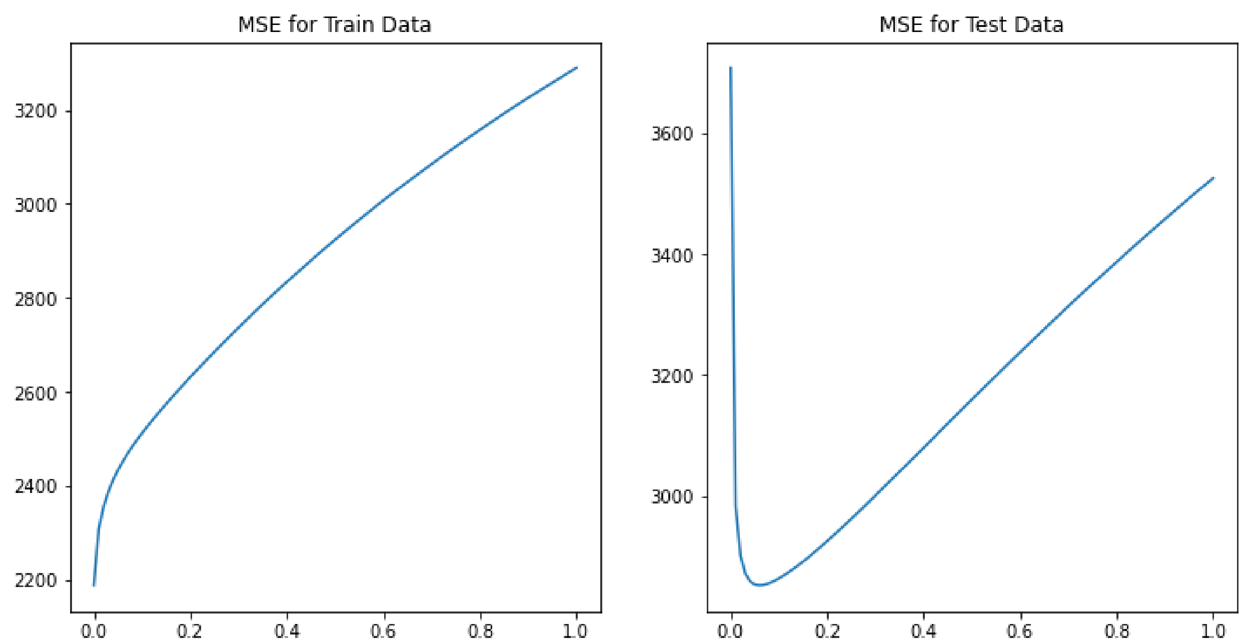
Train Data:
MSE without intercept 19099.446844570608
MSE with intercept 2187.160294930388

Test Data:
MSE without intercept 106775.36155260474
MSE with intercept 3707.840181718641

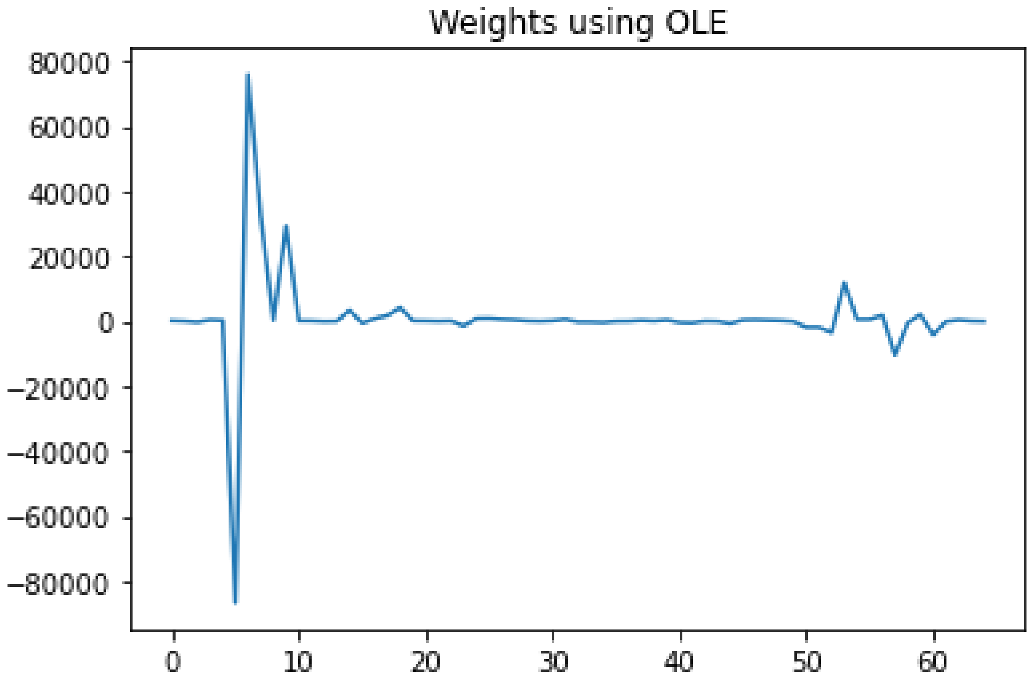
We can observe that the MSE values of train and test data with intercept is lower than that of without intercept. The line which fits the data has to pass through origin when intercept is not considered, so the error is high. Whereas when intercept is included for linear regression, the line gets more aligned with dataset and it need not pass through origin. That is why the error is less in this case. So, linear regression with intercept is better for both train and test data.

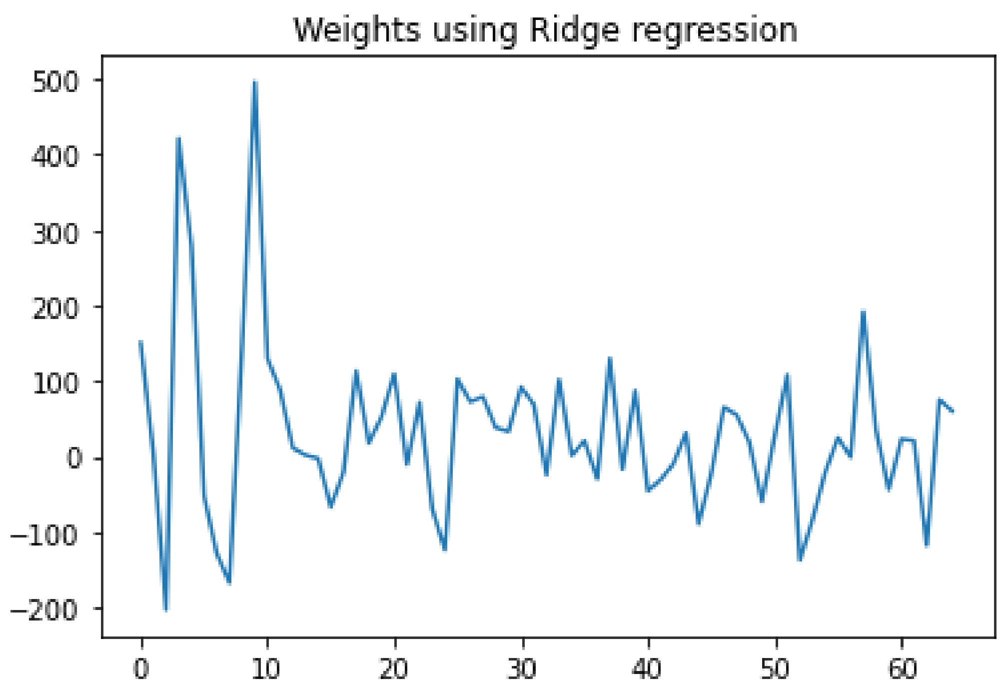
Problem 3: Experiment with Ridge Regression

MSE for training and test data using ridge regression



Relative magnitudes of weights learnt using OLE and ridge regression:





From the above graphs we can see that weights obtained using linear regression wit h intercept range from -80000 to 80000 and the weights obtained using ridge regres sion range from -200 to 500. By taking l2 norm of the difference of the weights we get 124519.268 which indicates that the regularization prevents the weights to be very large.

Compare the two approaches in terms of errors on train and test data:

Methods	Train MSE	Test MSE
Linear Regression without Intercept	19099.446	106775.361
Linear Regression with Intercept	2187.160	3707.840
Ridge Regression with optimal lambda(0.06)	2451.528	2851.330

From above table, we can observe ridge regression performs better than linear regression with and without intercept on test data.

Optimal value of lambda:

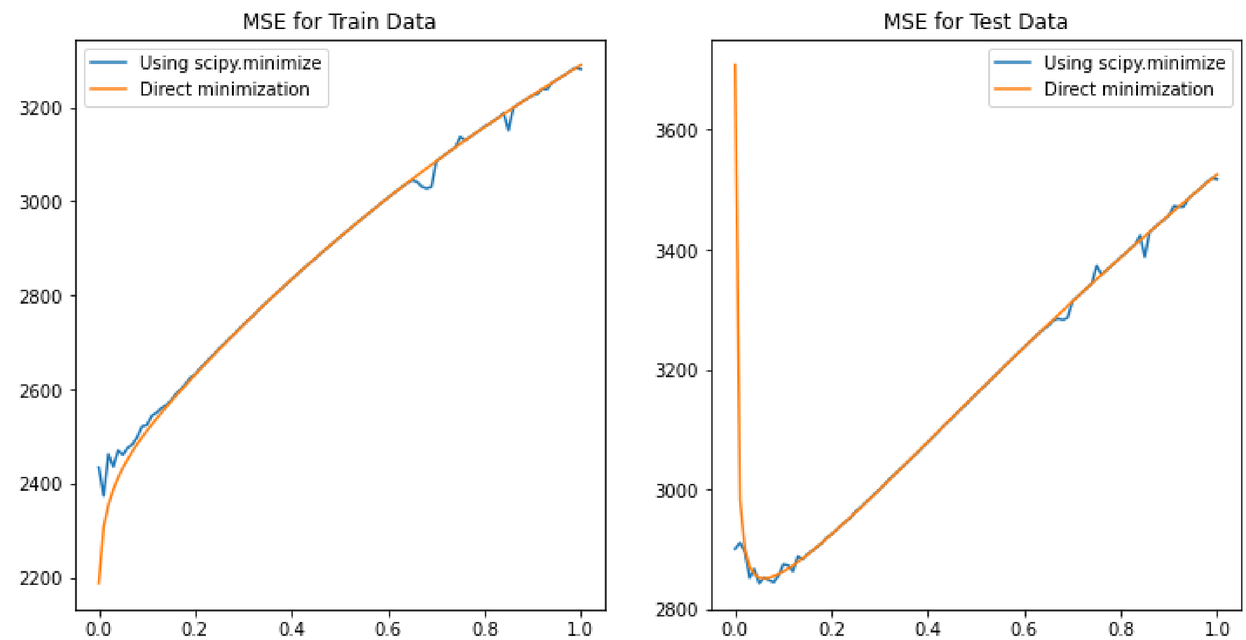
The error for the lambda value of 0.06 is lowest hence it is the optimal value. The MSE values for λ 0.06 are:

MSE for optimal value of lambda on train data 2451.52849064
MSE for optimal value of lambda on test data 2851.33021344

Problem 4: Using Gradient Descent for Ridge Regression Learning

Errors on train and test data obtained by using the gradient descent

For max_iter = 20:

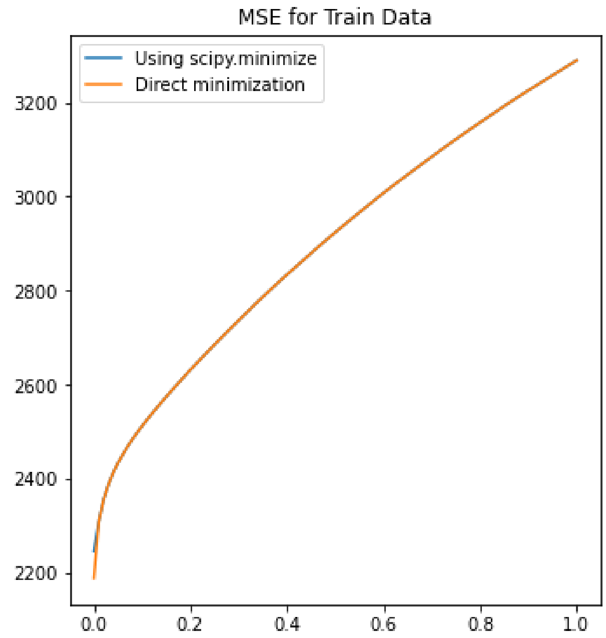


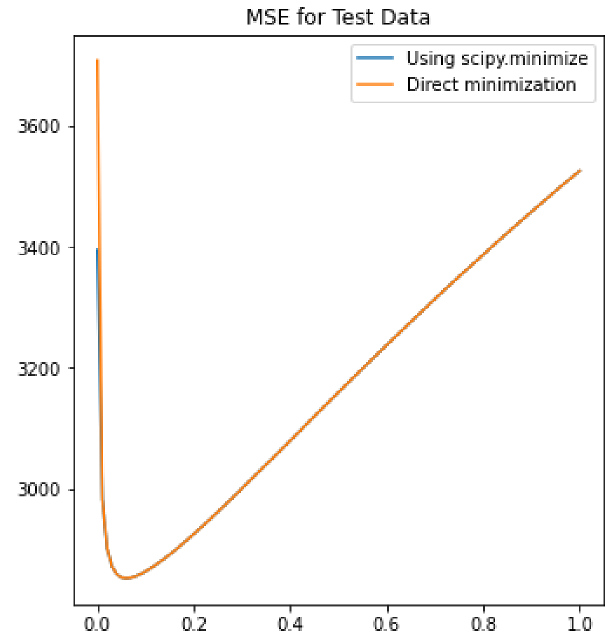
Compare with the results obtained in Problem 3:

Method	MSE on train data	MSE on test data
Ridge Regression	2451.528	2851.330
Ridge Regression using Gradient Descent(max_iter=100)	2451.528	2851.329

By comparing the results from above table, we can conclude that train and test MSE for ridge regression and ridge regression using gradient descent is similar.

For max_iter :100





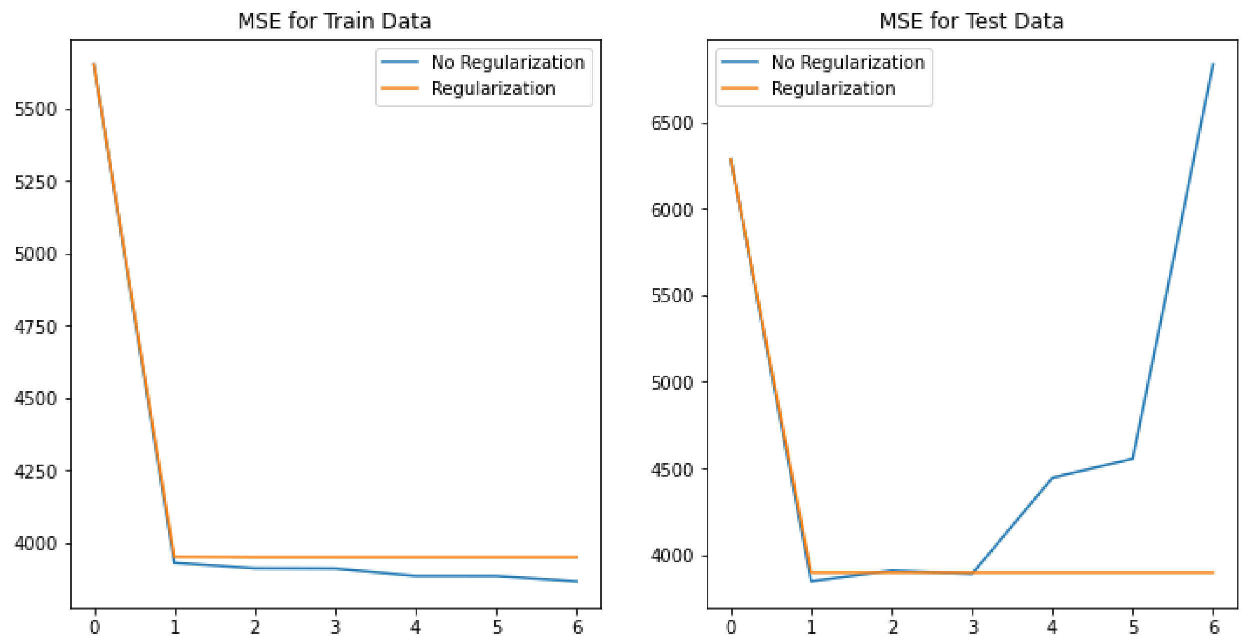
For lower iterations such as 20 the error fluctuates in case of gradient descent, but as we increase max_iter of minimize function to 100, the error curve of gradient descent is aligned with the error curve of normal ridge regression

Problem 5: Non-linear Regression

Errors on train and test data:

$\lambda = 0$ – No regularization

$\lambda = 0.06$ - regularization



No regularization:

When there is no regularization i.e., for lambda value 0, the model gets adjusted according to the training data and loses generalization. So, MSE value decreases on training data and increases on test data as we increase polynomial degree p.

Regularization:

When there is regularization i.e., for optimal value of lambda 0.06, the model is generalized for both train and test data. The curves become non-linear and MSE values starts decreasing and saturates for higher degree polynomials p.

Optimal value of p:

Lambda	Optimal p	MSE
0	1	3845.034730173414
0.06	4	3895.582668283526

Problem 6: Interpreting Results

Method	Train MSE	Test MSE
Linear Regression without intercept	19099.446	106775.361
Linear Regression with intercept	2187.160	3707.840
Ridge Regression	2451.528	2851.330
Ridge Regression with gradient descent	2451.528	2851.329
Non Linear regression (no regularization)	3866.883	3845.034
Non Linear regression(regularization)	3950.682	3895.582

The metric that can be used for best setting is lower Test MSE.

Final recommendations for predicting diabetes level:

As the ridge regression with gradient descent performs better with lowest MSE on test dataset, this approach can be used for predicting diabetes level to give the best accuracy.

Though the train error with ordinary linear regression is very less, the test error is high, which leads to overfitting. Whereas in ridge regression the difference between train and test error is less which means the model is generalized and controls overfitting. Also, using ridge regression with gradient descent, we can avoid inverse calculation on dot product of X transpose and X. This saves lot of computation time when working on relatively larger datasets.