CSE574 Fall 2021 Introduction to Machine Learning
Programming Assignment 1

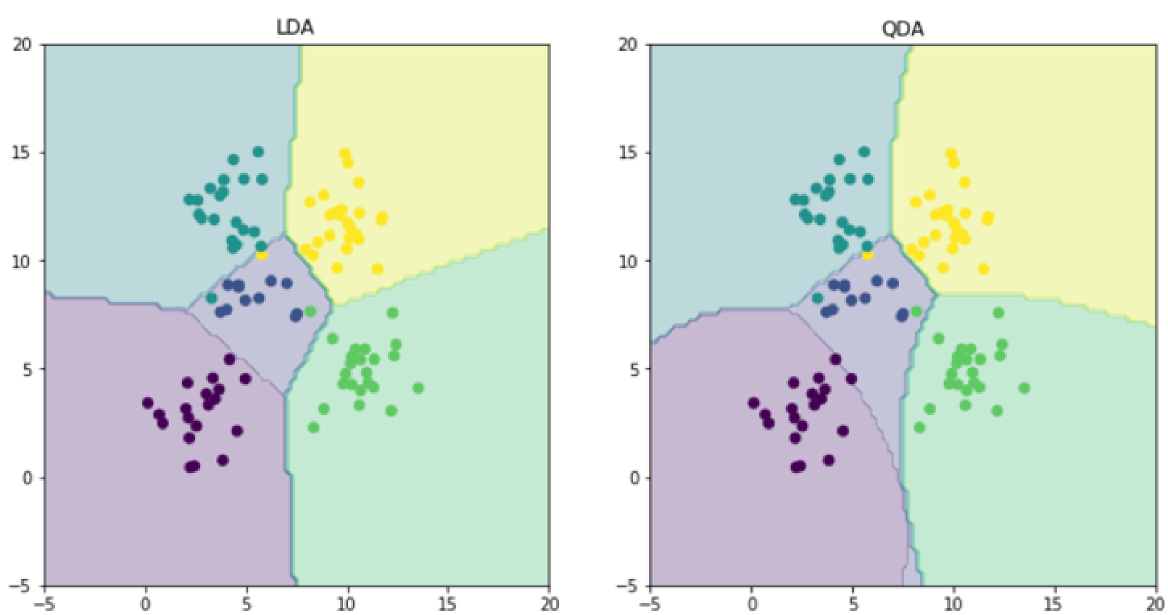
Classification and Regression

Group 26

NAME	UB PERSON ID:	UB IT NAME:
Akshaya Mohan	50419115	akshayam
Oviyaa Balamurugan	50418472	oviyaaba
Vaishnavi Ruppa Gangadharan	50418483	vruppaga

Problem 1: Experiment with Gaussian Discriminators

LDA Accuracy	97
QDA Accuracy	96



Both LDA and QDA are generative models which assume that class conditional densities are normally distributed. In both LDA and QDA, we train the model by learning the mean vector and covariance matrix. The difference is that, in QDA we learn a covariance matrix for each output class whereas in LDA we learn a single covariance matrix for the entire training data set irrespective of the class. Therefore, LDA relaxes the need for a lot of training data compared to QDA and there is also an absence of the quadratic term while calculating the PDF of the LDA. Due to this, the discriminating boundaries are non linear in the case of QDA and linear in the case of LDA.

Problem 2: Experiment with Linear Regression

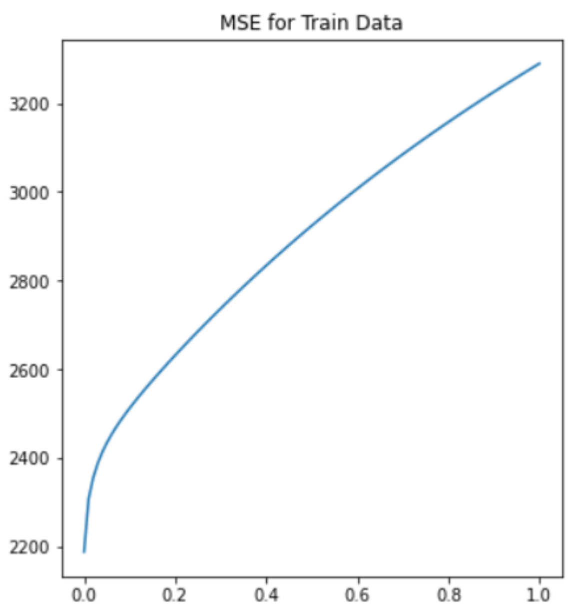
MSE for training data without intercept	19099.44684457
MSE for training data with intercept	2187.16029493
MSE for test data without intercept	106775.36156036
MSE for test data with intercept	3707.84018165

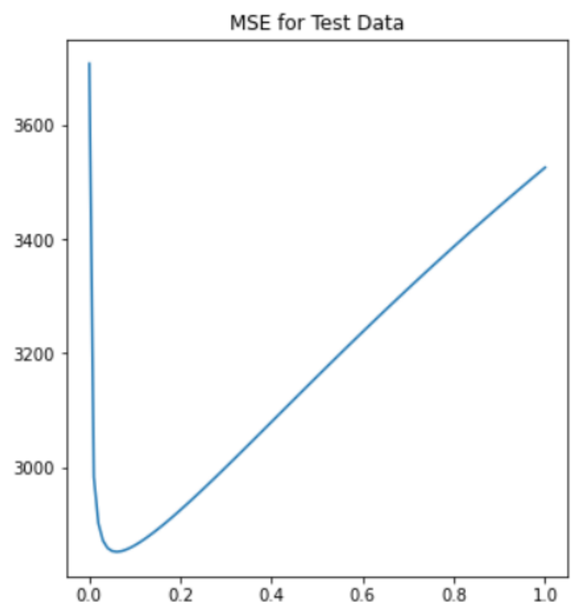
We add an intercept/bias term to the input data so that the line learnt through the linear regression does not have to pass through the origin. This reduces the Mean Squared Error and we get better results. Therefore, adding intercept to the input is the better approach.

Problem 3: Experiment with Ridge Regression

λ	MSE for training data	MSE for test data
0	2187.16029493	3707.84018165
0.01	2306.83221793	2982.44611971
0.02	2354.07134393	2900.97358708
0.03	2386.7801631	2870.94158888
0.04	2412.119043	2858.00040957
0.05	2433.1744367	2852.66573517

0.06	2451.52849064	2851.33021344
0.07	2468.07755253	2852.34999406
	. . .	
1	3289.7612813	3525.39455263



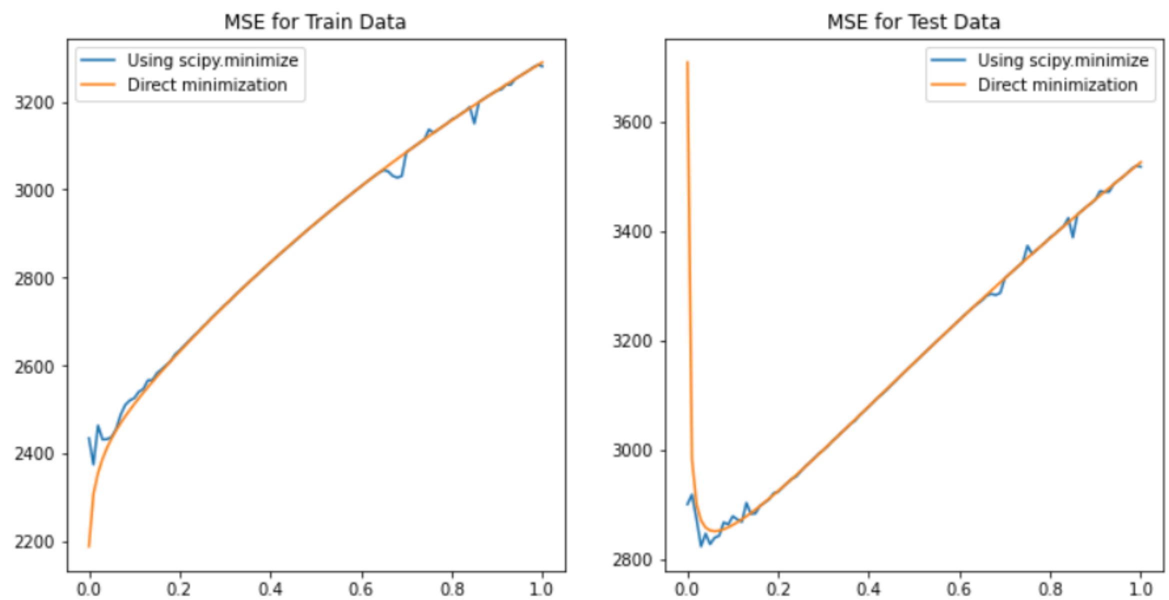


While comparing the relative magnitude of weights from both OLE and ridge regression, we observed that the weights learnt using OLE vary between the range -111.67740 and 203.3144. The weights learnt using Ridge Regression vary between the range -86639.457 and 75914.467. The weights are learnt from training data with intercept.

	OLE	**Ridge Regression (at optimal λ)**
MSE for training data with intercept	2187.16029493	2187.16029493 (λ=0)
MSE for test data with intercept	3707.84018165	2851.33021344 (λ=0.06)

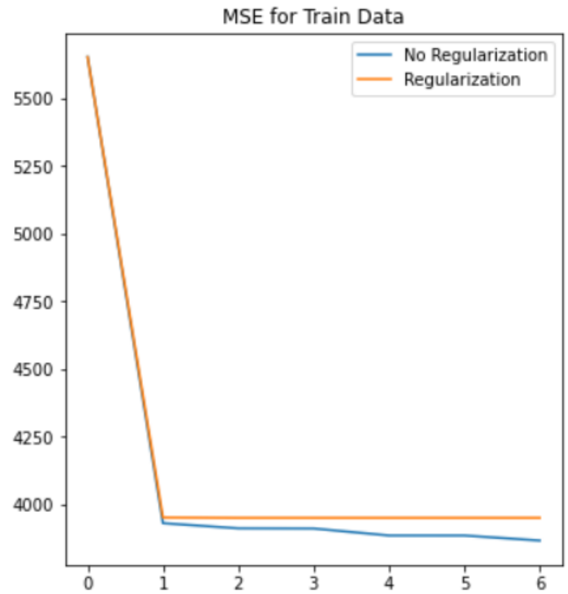
As seen from the table above, MSE for training data for both linear regression and ridge regression are the same.
The optimal λ is **0** for training data and **0.06** for test data as the MSE is lowest at those values of λ.

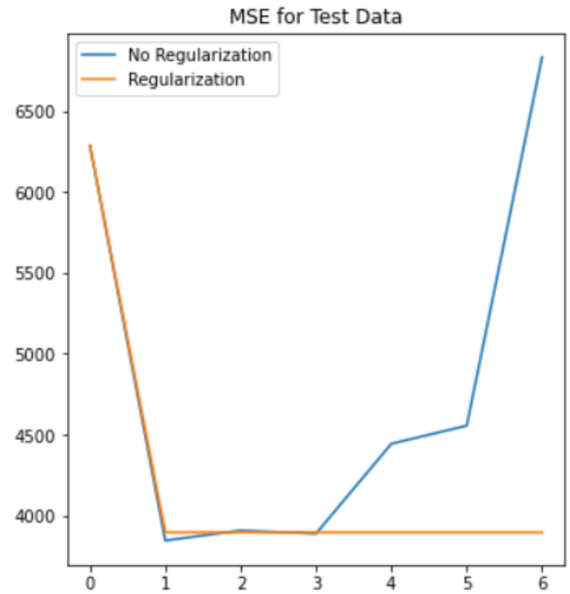
Problem 4: Using Gradient Descent for Ridge Regression Learning



As seen from the plot, orange line indicates weights learnt for ridge regression and blue line indicates weights learnt for ridge regression using gradient descent approach. It is also noted that the number of iterations used for the gradient descent approach is 20.

Problem 5: Non-linear Regression





p	MSE for test data with no regularization	MSE for test data with regularization
0	6286.40479168	6286.88196694
1	3845.03473017	3895.85646447
2	3907.12809911	3895.58405594
3	3887.97553824	3895.58271592
4	4443.32789181	3895.58266828
5	4554.83037743	3895.5826687
6	6833.45914872	3895.58266872

As seen from the table above, the optimal value of p is **1** with no regularization and it is **4** with regularization as the MSE is lowest at those values of p.

Problem 6: Interpreting Results

So far, we have used the same data set to make predictions using different regression approaches such as linear regression, ridge regression, ridge regression with gradient descent and finally non-linear regression. To find out which is the best approach for the given data set, we compare the lowest MSE values obtained using each of these approaches. The values reported all use data with intercept.

	Lowest MSE for training data	**Lowest MSE for test data**
Linear Regression	2187.16029493	3707.84018165
Ridge Regression (at optimal λ)	2187.16029493	2851.33021344
Ridge Regression with gradient descent	2406.30544593	2832.39053852
Non Linear Regression with no regularization	3866.88344945	3845.03473017
Non Linear Regression with regularization	3950.68233514	3895.58266828

From the table above, it is clear that Ridge regression is the best approach. Particularly, ridge regression with gradient descent is the recommended approach for predicting diabetes level using the input features. This recommendation is given based on two main reasons,

- Ridge regression with weights learnt using gradient descent approach gives the lowest MSE values and due to this, predictions will be better
- Gradient descent approach also avoids the inverse calculation required for learning weights which can be particularly tricky when there is singularity involved in the input dataset