CSE574 Introduction to Machine Learning

Programming Assignment 1

Classification and Regression

Haowei Zhou

Problem 1:

LDA Accuracy = 0.97

QDA Accuracy = 0.96

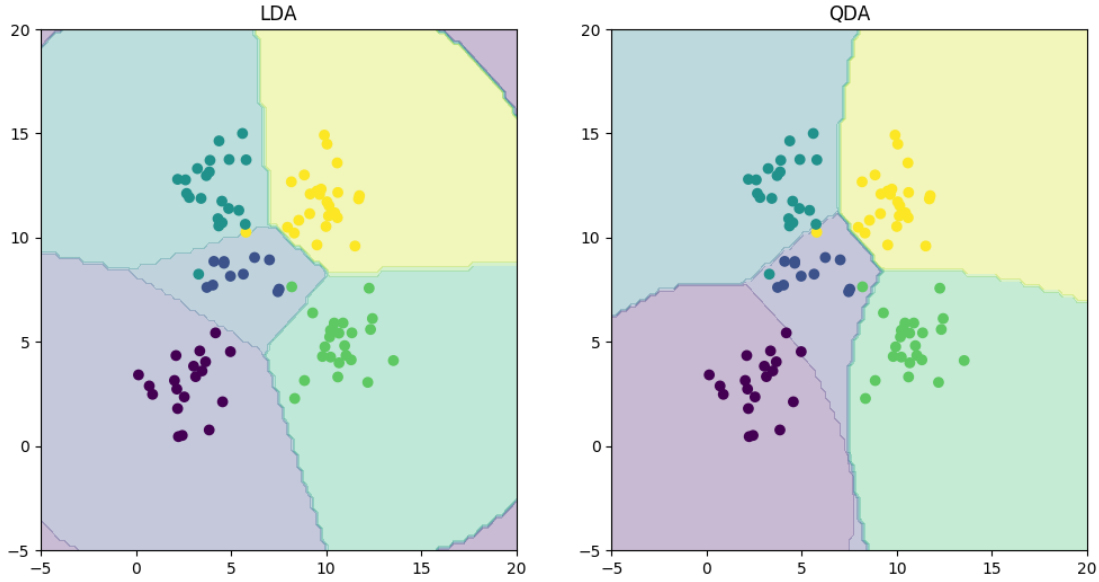

Figure 1.

The difference of the boundary between LDA and QDA is the boundary of LDA is linear and the boundary of QDA is quadratic. The reason is the covariance matrix of LDA is the same for all classes, but the covariance matrix of QDA is calculated by each classes. Because LDA can only learn linear boundary but QDA can learn quadratic boundary, so we can conclude QDA is more flexible than LDA.

Problem 2:

	Test data	Train data
With intercept	3707.84018193	2187.16029493
Without intercept	106775.36153755	19099.44684457

Table 1. MSE of test data and train data

According to the test data with and without intercept, we can see the model with intercept performs better than the model without intercept.

Problem 3:

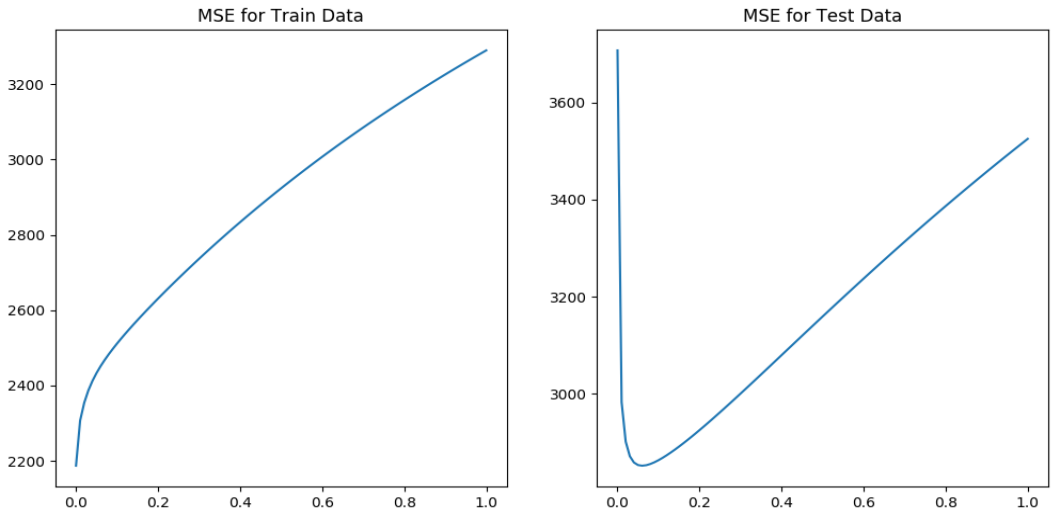

Figure 2.

For training data, the minimum of MSE is lambda equals to 0, which means there is no regularization. However, for test data, the MSE get its minimum when lambda equals to 0.06. The MSE is greater when lambda is greater or smaller than 0.06. When lambda is smaller than 0.06, the prior regularization has limited influence on test data, so the training data is overfitting so MSE is very large. On the other hand, when lambda greater than 0.06, the prior regularization has dominant influence on test data, so the MSE linearly increases just like train data.

Problem 4:

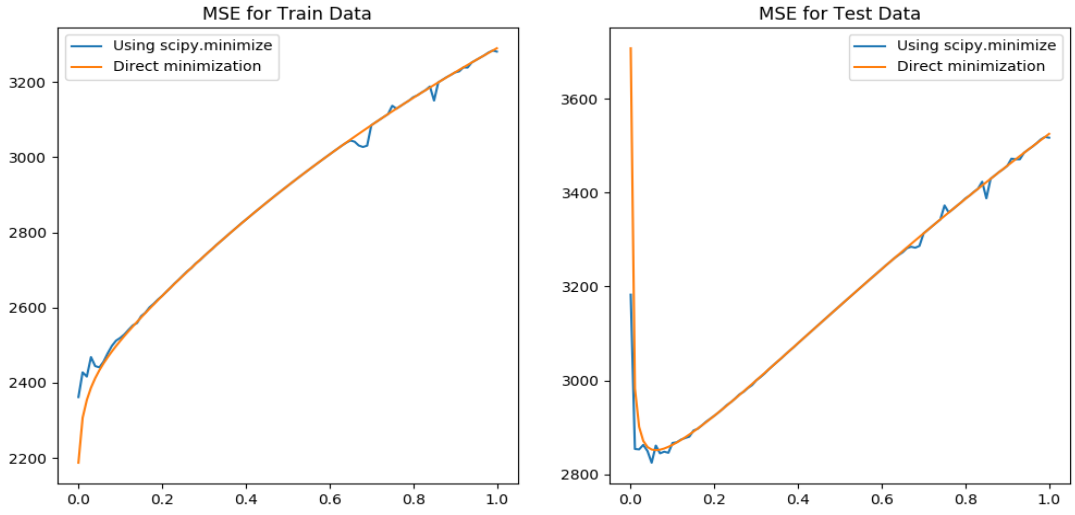

Figure 3.

Comparing the result we get in problem 3, the MSE is roughly overlapping in terms of gradient decent and direct minimization. And we can see if we increase the number of iteration, the MSE looks more smooth and more close to the result of direct minimization.

Problem 5:

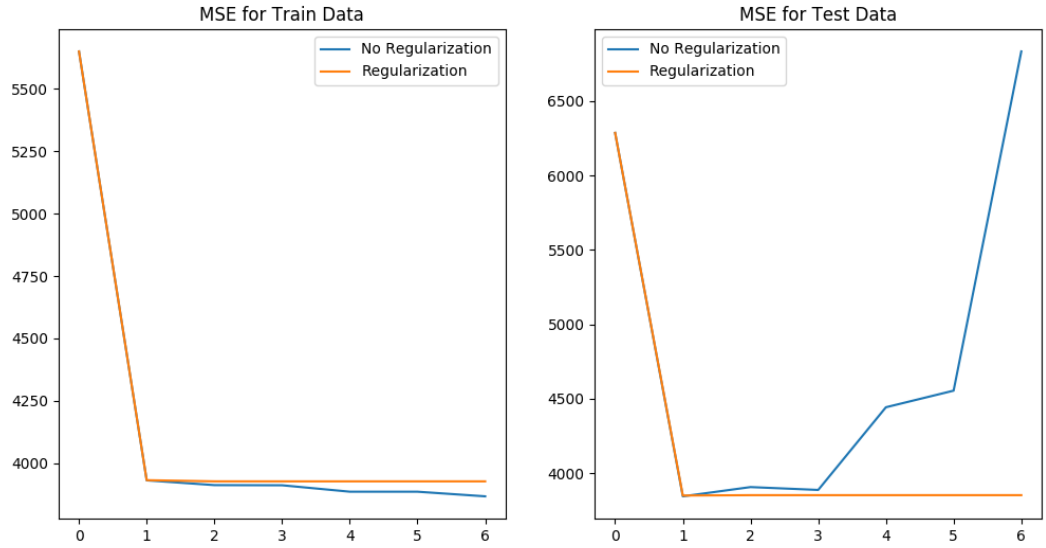

Figure 4.

When p equals to 1 in terms of MSE when lambda equals to 0.06. Because if p is greater than 1, the model has better performance on train data, however, it might cause overfitting, so the MSE of test data in terms of no regularization is rapidly increase when p greater than 1. If we use regularization, the prior makes the model more stable.

Problem 6:

According to different kinds of regression methods, the model of ridge regression has the best performance for predicting the data. The MSE is smallest when lambda equals to 0.06 and p equals to 1.