CSE574 Introduction to Machine Learning

Programming Assignment 1

Project Report: Classification and Regression (PA1)

Submitted by:

Nimisha Philip Rodrigues 5024 8016

Manasi Samir Tamboli 5024 6653

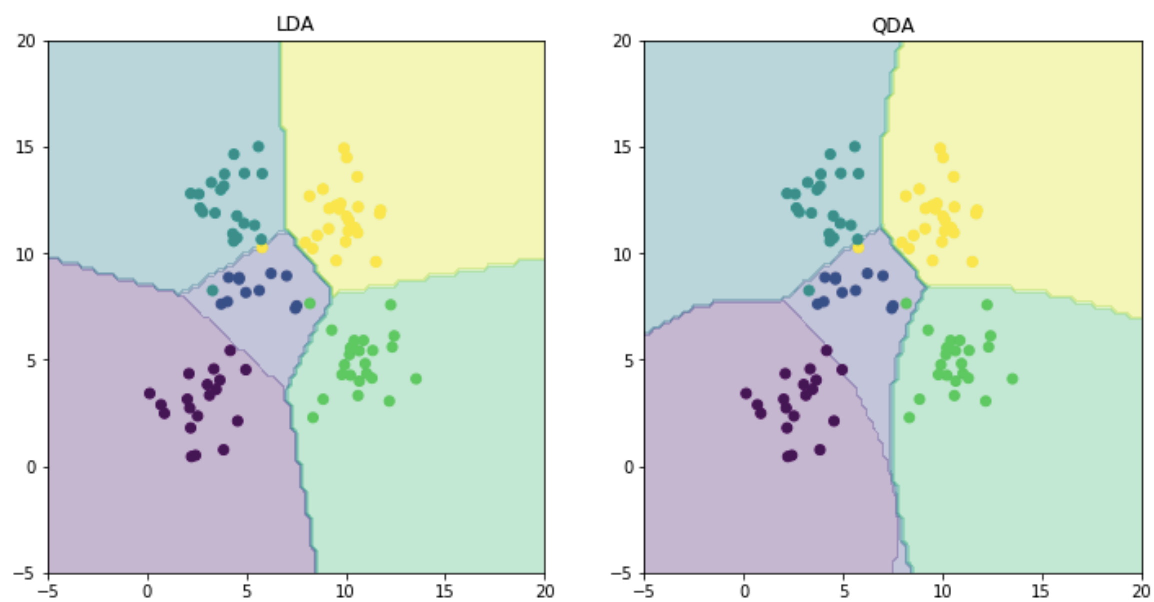
Problem 1: Experiment with Gaussian Discriminators

Question: Train both methods using the sample training data (sample train). Report the accuracy of LDA and QDA on the provided test data set (sample test). Also, plot the discriminating boundary for linear and quadratic discriminators. The code to plot the boundaries is already provided in the base code. Explain why there is a difference in the two boundaries.

Result:

a) Accuracy:
   LDA = 97
   QDA = 96

b) Plotting Graphs:



c) Explain the difference in two boundaries:

In LDA, even though the classes are different the covariance matrix is same, whereas for QDA we use different covariance matrix for each class. This difference gives kind of linear boundary for LDA and kind of non- linear for QDA.

Problem 2: Experiment with Linear Regression

Question:

Calculate and report the MSE for training and test data for two cases: first, without using an intercept (or bias) term, and second with using an intercept. Which one is better?

Result:

a. Observation:

Data Type	Mean Squared Loss (MSE) without using intercept	Mean Squared Loss (MSE) with intercept	% Improvement
Train Data	19099.4468446	2187.16029493	88.54%
Test Data	106775.361558	3707.84018132	96.5%

b. Which one is better?

The MSE with intercept is better than the MSE without intercept. The MSE without intercept returns a comparatively larger value for both train data and test data as compared to the MSE with intercept. We can say MSE with intercept provides better accuracy since the loss values are small. The graph of MSE without intercept passes through origin and hence reduces the accuracy as opposed to MSE with inclusion of an intercept.

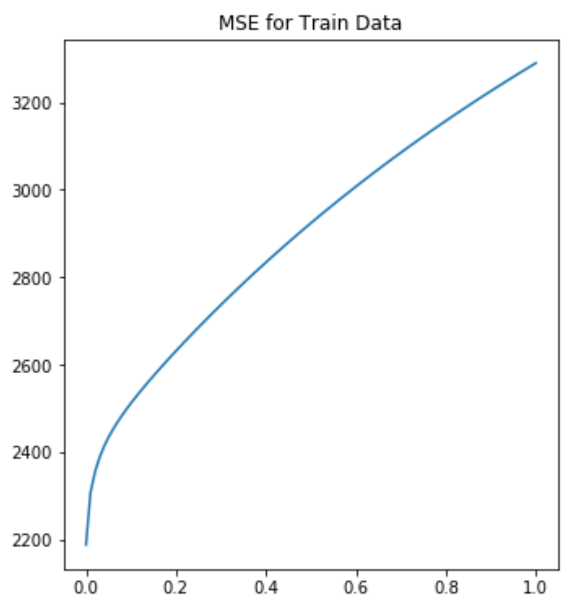
Problem 3: Experiment with Ridge Regression

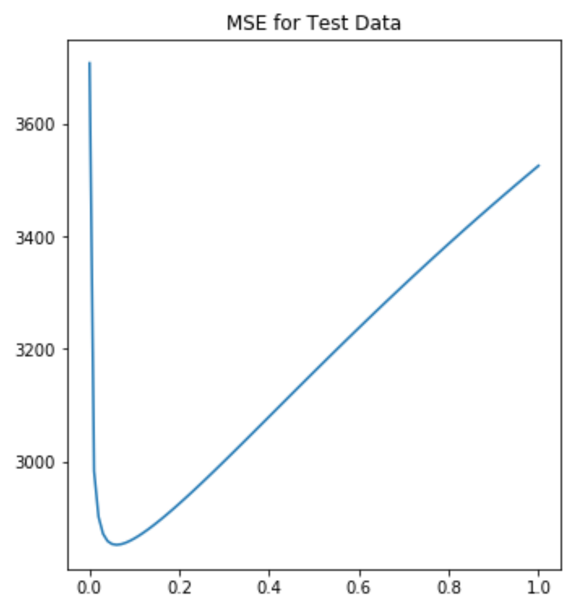
Question:

Calculate and report the MSE for training and test data using ridge regression parameters using the testOLERegression function that you implemented in Problem 2. Use data with intercept. Plot the errors on train and test data for different values of λ. Vary λ from 0 (no regularization) to 1 in steps of 0.01. Compare the relative magnitudes of weights learnt using OLE (Problem 2) and weights learnt using ridge regression. Compare the two approaches in terms of errors on train and test data. What is the optimal value for λ and why?

Results:

a. MSE for training and test data using ridge regression parameters

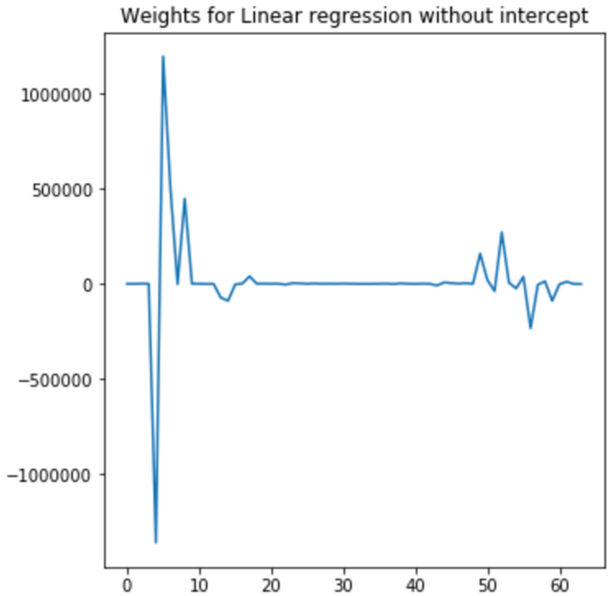


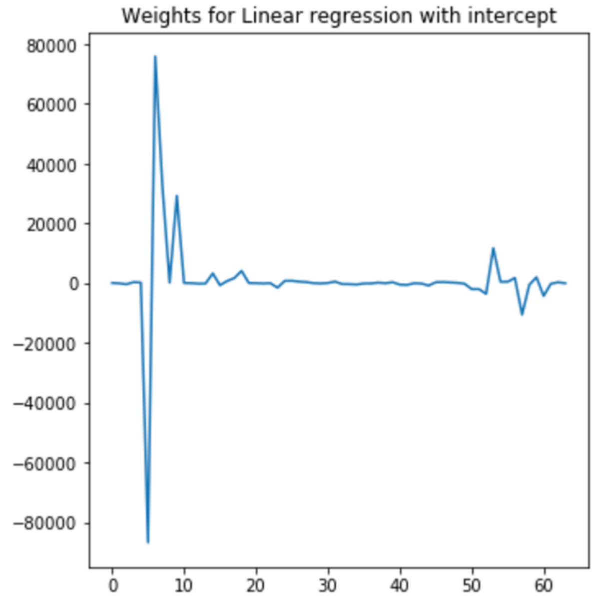


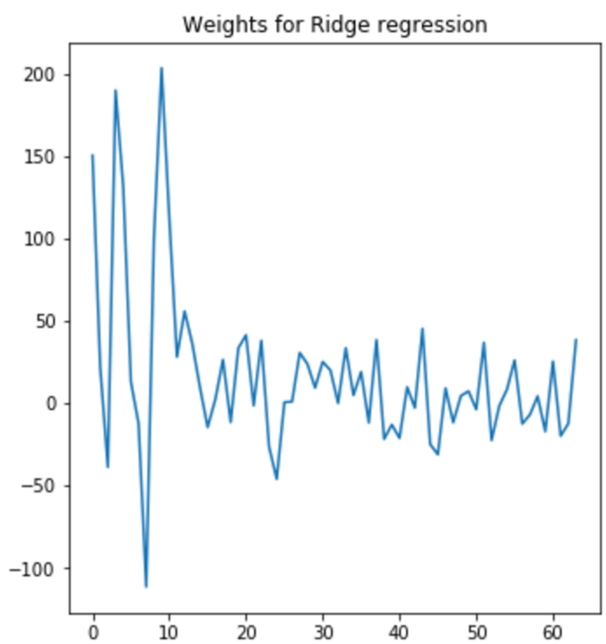
Observation:
As we go on increasing the value of lambda λ we get larger values for the MSE for both training and test data. So MSE is directly proportional to the lambda value.

Lambda	**MSE for train data**	**MSE for test data**
0.0	**2187.16**	3707.84
0.01	2306.83	2982.45
0.02	2354.07	2900.97
0.03	2386.78	2870.94
0.04	2412.12	2858.0
0.05	2433.17	2852.67
0.06	2451.53	**2851.33**
0.07	2468.08	2852.35
0.08	2483.37	2854.88
...	...	...
0.98	3277.26	3512.04
0.99	3283.53	3518.73
1.0	3289.76	3525.39

b. Relative magnitudes of weights learnt using OLE and weights learnt using ridge regression.







Observation:

Range of weights for ridge regression varies from -100 to +200.

Range of weights learnt using OLE without intercept varies from -10,00,000 to + 10,00,000.
Range of weights learnt using OLE with intercept varies from -80,000 to +80,000.

a. **Compare the two approaches in terms of errors on train and test data.**

	Lambda	**MSE for Train Data**	**MSE for Test Data**
OLE	Without intercept	19099.44	106775.36
	With intercept	2187.16	3707.84
Ridge Regression	0	2187.16	3707.84
	0.06	2451.52	2851.33

Observation:
Ridge Regression gives best results for test data when lambda value is set to 0.06 with least error values and for train data when lambda set to 0.0. Also, we get same values of MSE for both train and test data for Ridge Regression using lambda as 0.0 and when we perform OLE with intercept. OLE without intercept returns a fairly large value of MSE for both train and test data.

b. **What is the optimal value for λ and why?**

Optimal value of lambda depends on the least value we get for MSE. So the optimal value for λ is 0.0 for train data and 0.06 for test data.

Problem 4: Using Gradient Descent for Ridge Regression Learning

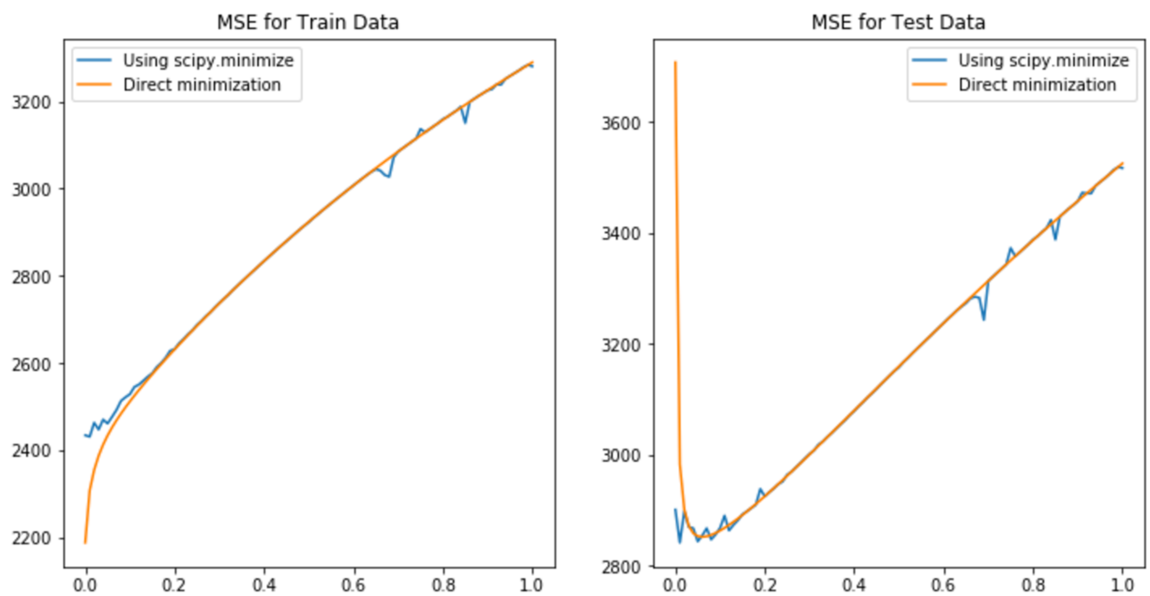
Question:

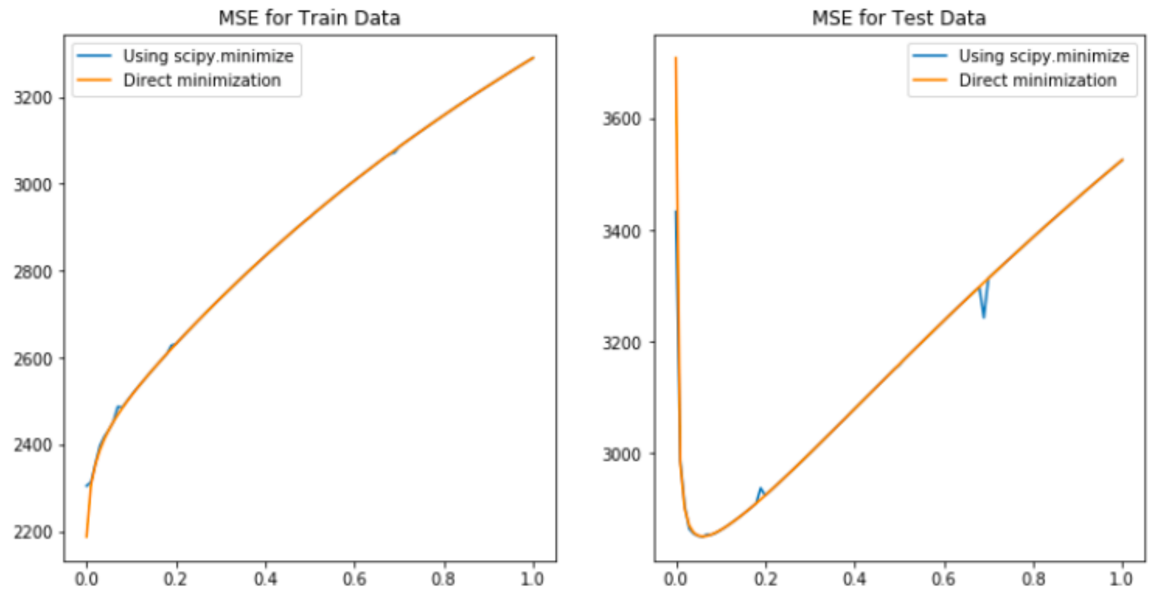
Plot the errors on train and test data obtained by using the gradient descent based learning by varying the regularization parameter λ. Compare with the results obtained in Problem 3.

Results:

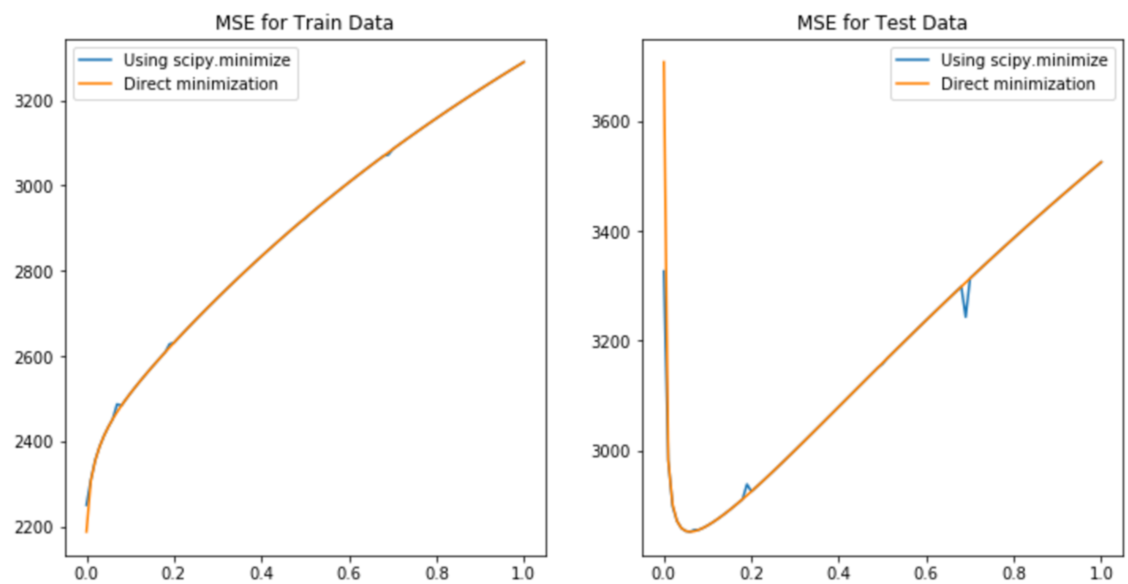
For iteration = 20

For iteration = 50

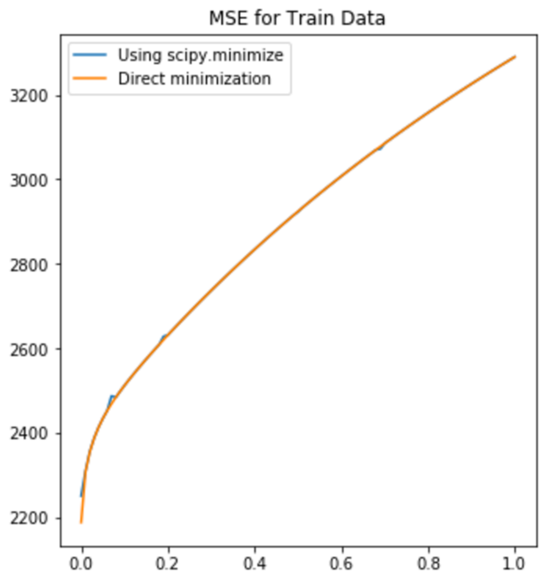


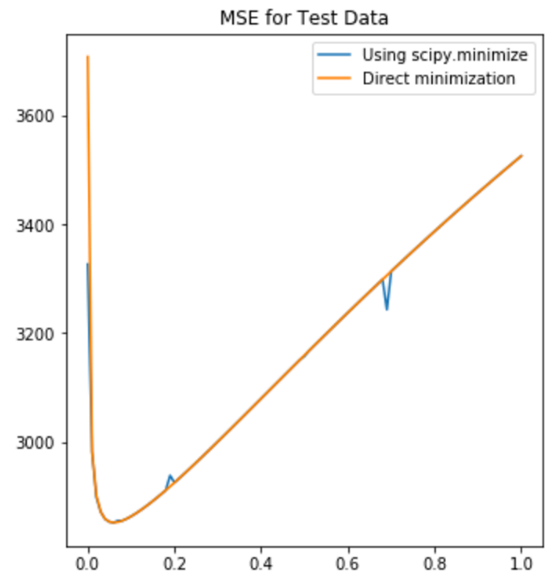


For iteration = 100



For iteration = 150





Observation:

The errors returned by using scipy.minimize at lower iterations such as 20 and 50 deviate a little from the results obtained by direct minimization. The scipy.minimize learns more ass we increase the number of iterations, and we get results almost similar to the ones returned by direct minimization.

Problem 5: Non-linear Regression

Question:

Using the λ = 0 and the optimal value of λ found in Problem 3, train ridge regression weights using the non-linear mapping of the data. Vary p from 0 to 6. Note that p = 0 means using a horizontal line as the regression line, p = 1 is the same as linear ridge regression. Compute the errors on train and test data. Compare the results for both values of λ. What is the optimal value of p in terms of test error in each setting? Plot the curve for the optimal value of p for both values of λ and compare.

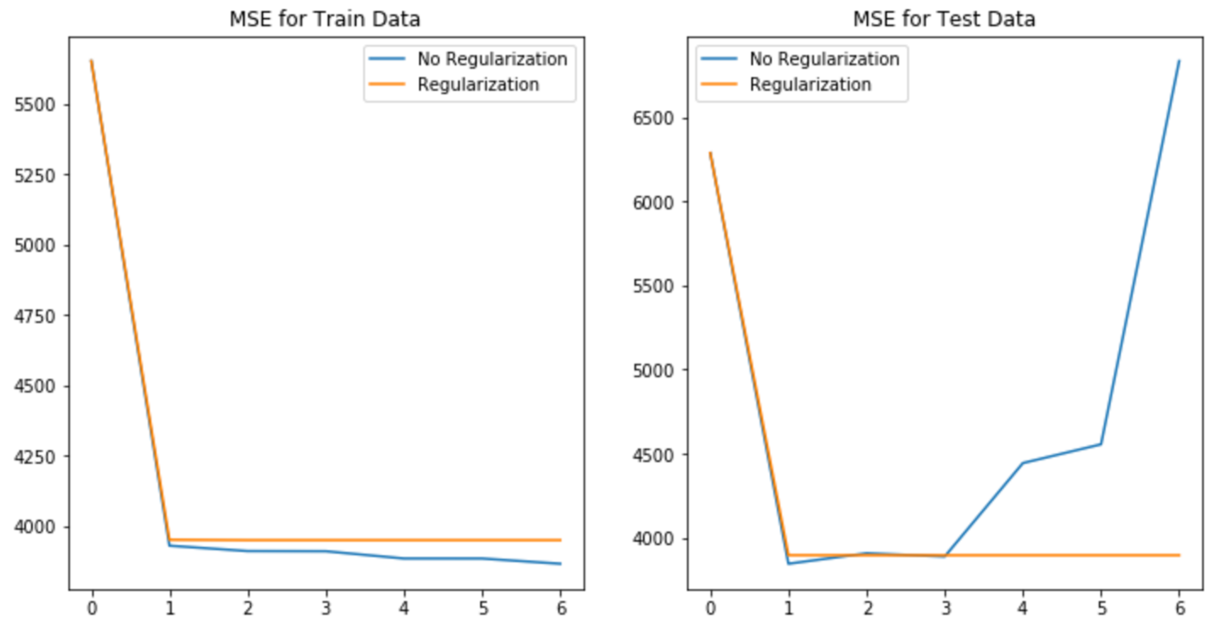
a. Compare the results for both values of λ

p	MSE (No regularization)	MSE (regularization)
0	6286.4	6286.88
1	**3845.03**	3895.86
2	3907.13	3895.58
3	3887.98	3895.58
4	4443.33	**3895.58**
5	4554.83	3895.58
6	6833.46	3895.58

b. What is the optimal value of p in terms of test error in each setting?

Optimal value of p (no regularization) is **1**
Optimal value of p (with regularization) is **4**

c. Plot the curve for the optimal value of p for both values of λ and compare.



Problem 6: Interpreting Results

Question:
Compare the various approaches in terms of training and testing error. What metric should be used to choose the best setting?

Result:

- Linear OLE Regression with intercept:
  In this the model passes through origin, thus results are inaccurate.

- Linear OLE Regression without intercept:
  Since there is an intercept here, model passes through the data and give better results.

- Ridge Regression:
  Penalties are added to the magnitude of coefficients.

- Ridge Regression with Gradient Descent:
  Log likelihood is maximized so that calculation of inverse matrix is avoided.

- Non- linear Regression with p=1:
  In this we give higher order polynomials as input.

- Non- linear Regression with p=4:
  In this also we give higher order polynomials as input.