1st Project - Regression

Role	Participation	Name
Team Member	1.0	Daehoon Gwak (dps5313@psu.edu)
Team Member	1.0	Jaehoon Ha (jxh5648@psu.edu)
Team Member	1.0	Daniel Jung (dqj5182@psu.edu)

Problem 1: Experiment with Linear Regression

1. For both LinearRegressionModel and learnOLERegression, calculate and report the MSEs for training and test data for two cases: first, without using an intercept (or bias) term, and second with using an intercept

   : By the calculation from LinearRegressionModel and learnOLERegression function, we got 4 different cases: Sklearn MSE without intercept, Sklearn MSE with intercept, OLE MSE without intercept, OLE MSE with intercept. Each of these values has been calculated with our functions of LinearRegressionModel and learnOLERegression and given the main script. The outputted values are (1) Sklearn MSE without intercept: **106775.36145123**, (2) Sklearn MSE with intercept: **3707.84018096**, (3) OLE MSE without intercept: **106775.36155731**, (4) OLE MSE with intercept: **3707.84018145**.

2. Are the results outputted by LinearRegressionModel and learnOLERegression similar?

   : We can see from the resulted data that each "without intercept" data and "with intercept" data from Sklearn MSE and OLE MSE are both very similar to each other. There doesn't seem to be much distinction among such values.

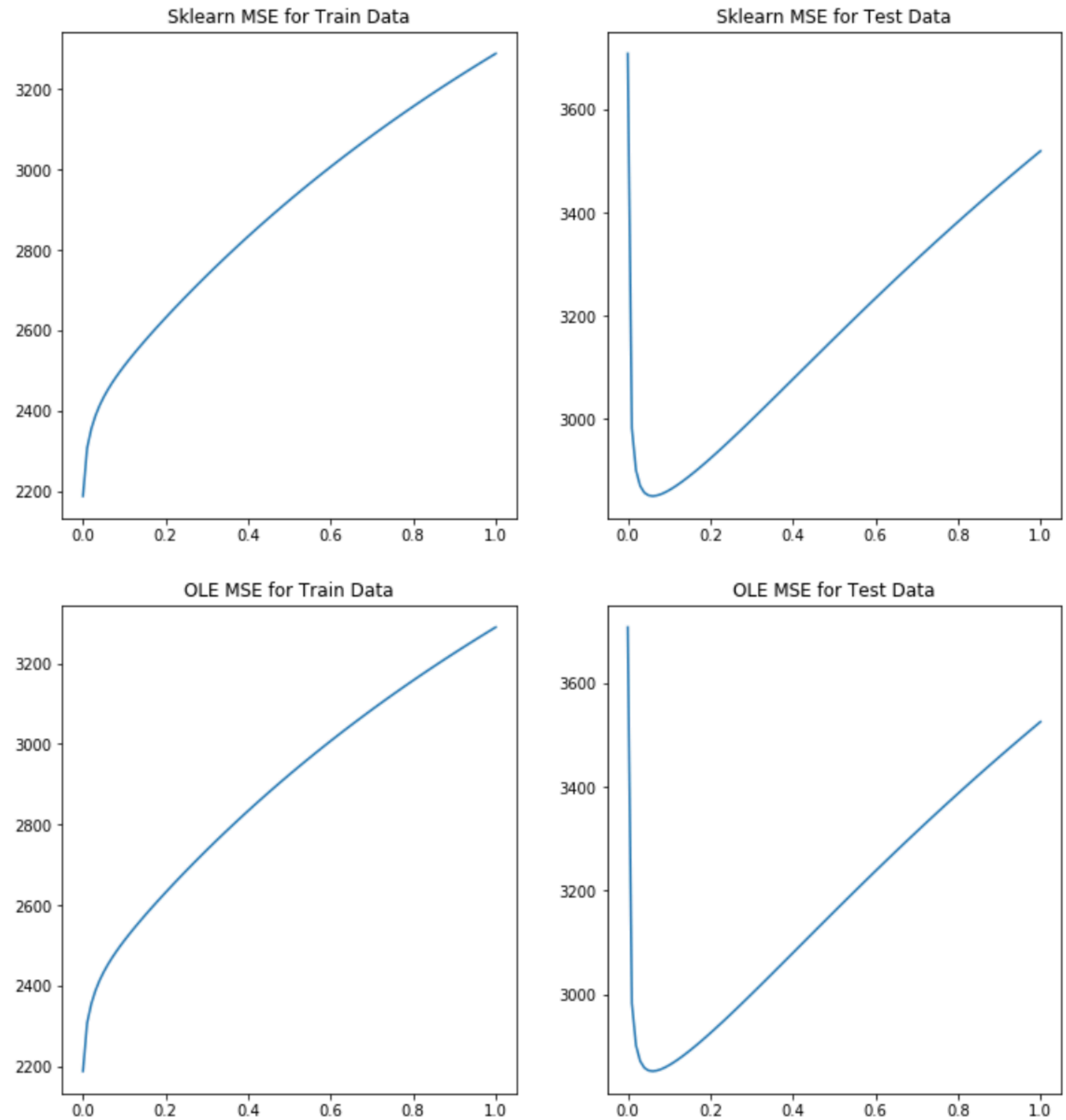
3. According to these testing results, do you suggest using an intercept (or bias) term for linear regression problems?

   : Using such a result, we would suggest using "with intercept" term for linear regression problems, since MSE value for both Sklearn method and OLE method were much smaller than when using the "without intercept" term.

Problem 2: Experiment with Ridge Regression

1. Plot the errors on train and test data for different values of α. Vary α from 0 (no regularization) to 1 in steps of 0.01

—



<Figure 1>

2. Are the smallest MSEs outputted by learnRidgeRegression and RidgeRegressionModel similar? What are the optimal values of α's to achieve the smallest testing MSEs?

: We can visually check the smallest MSEs for both Sklearn and OLE methods from the above 'Figure 1'. It seems very likely that they are very similar and almost identical. Also, we can double-check the similarity by checking the smallest MSE outputted by learnRidgeRegression: 2850.8958561367936 and MSE from RidgeRegressionModel: 2851.3302134438477. So, we can confirm that they are very similar.

We can also find from the output value that the optimal alpha value to achieve the smallest testing MSEs are both 0.06.

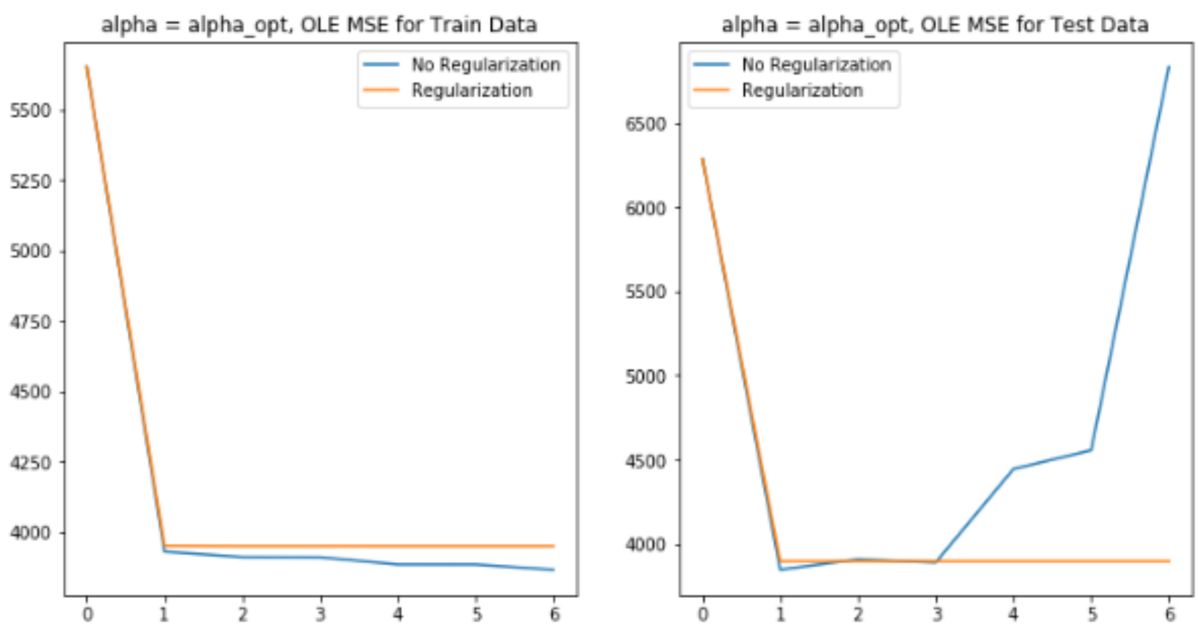
3. Compare the four approaches (two in Problem 1 and two in this problem) in terms of errors on train and test data.

: We can now compare four approaches from four different functions. When we try to find the smallest MSE and corresponding alpha, the value was almost the same when whether we use Sklearn or OLE function. Also, we can figure out the MSE value was smaller when we used test data than when we used training data. Overall, the best approach to get the minimal value of error among the four approaches seems to be "Ridge Regression using Sklearn MSE" which can reach MSE as small as 2850.

Problem 3: Non-linear Regression

1. Compute the errors on train and test data for different p's. Compare the results for both values of α (α = 0 means No Regularization, and α = alpha opt means Regularization).

   : Through our functions and pre-given main script, we could get the result for both values of α for different p values. The graph below shows the testing error for both No regularization (α = 0) and Regularization (α = alpha opt). The No regularization is in color "blue" and Regularization (α = alpha opt) is in color "orange". If we compare the result for both values of α, we can see that for the training data, there isn't much difference in OLE MSE value. However, for the testing data, there is significance difference of OLE MSE value for each value of α.



<Figure 2>

   a. Training errors of linear models without using the regularization term:

      p = 0 training error = 5650.710538897616

      p = 1 training error = 3930.9154073159016

p = 2 training error = 3911.8396712049557

p = 3 training error = 3911.1886649314497

p = 4 training error = 3885.473068112271

p = 5 training error = 3885.407157397081

p = 6 training error = 3866.88344944605

b. Training errors of linear models using the regularization term:

p = 0 training error = 5650.711907032115

p = 1 training error = 3951.8391235601057

p = 2 training error = 3950.6873123755195

p = 3 training error = 3950.682531518713

p = 4 training error = 3950.682336795369

p = 5 training error = 3950.6823351770195

p = 6 training error = 3950.6823351427815

c. Testing errors of linear models without using the regularization term:

p = 0 testing error = 6286.404791680896

p = 1 testing error = 3845.034730173414

p = 2 testing error = 3907.1280991079375

p = 3 testing error = 3887.9755382360127

p = 4 testing error = 4443.327891813304

p = 5 testing error = 4554.830377434741

p = 6 testing error = 6833.459148719205

d. Testing errors of linear models using the regularization term:

p = 0 testing error = 6286.881966941448

p = 1 testing error = 3895.8564644739613

p = 2 testing error = 3895.5840559389176

p = 3 testing error = 3895.5827159230994

p = 4 testing error = 3895.582668283526

p = 5 testing error = 3895.5826687044228

p = 6 testing error = 3895.582668719096

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

: The optimal values of p in terms of test error in each set are 1 when there is **no regularization** with testing error is **3845.034730173414**, and 4, when there is a **regularization** term with testing error, is **3895.582668283526**. So, it is better to use the regression without regularization in this case.

3. Plot the curve for the optimal value of p for both values of α and compare.

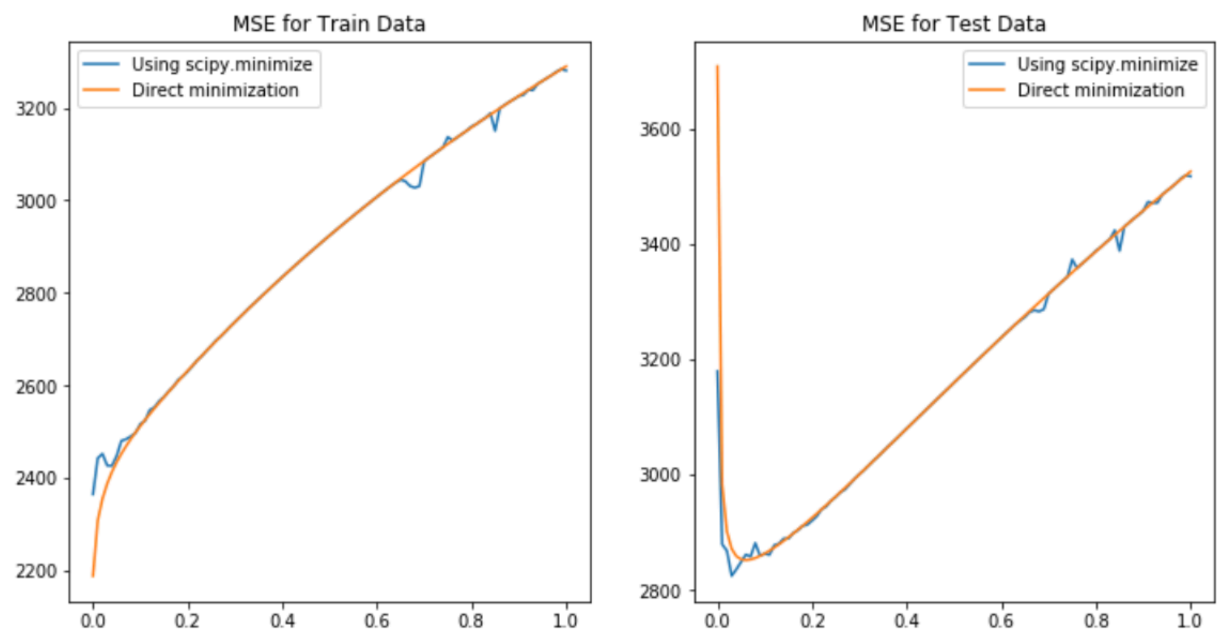
: As Figure 2 shows, for the training data, the No Regularization's MSE decreases after p=1 compared to the Regularization's MSE. However, In the testing data, the No Regularization's MSE increases compared to the Regularization's MSE after p=1, as the optimal values of p increases. The plot also shows that the optimal value of p = 1 for the 'No Regularization' and p = 4 for the 'Regularization' line.

Problem 4: Interpreting Results

There are three main ways for approaching the problem of predicting diabetes levels using the input features. The first approach was using the LinearRegressionModel and learnOLERegression method with and without intercept. In this approach, the most efficient approach to minimize MSE was “Sklearn MSE with intercept” which had an MSE value of **3707.84018096**. The second approach was using Ridge Regression. We were able to calculate and compare both Sklearn MSE and OLE MSE using the regression model. This model allowed us to reduce the MSE value to just **2850.8958561367936** with $\alpha = 0.06$. The third approach was using Non-linear Regression. We have used both the α value of 0 and α of optimal value found in the second approach. With this method, we could get the lowest testing error value of **3845.034730173414**.

According to our result error values from three different approaches, it is clear that the second approach with **Ridge Regression** was the best method of minimizing the error value for our data set.

Problem 5: Using Gradient Descent for Ridge Regression Learning



<Figure 3>

We now compare MSE for Train and Test data by varying the regularization parameter α. As 'Figure 3' shows, the output of the gradient descent-based learning and output from problem 2 are similar. However, MSE is even smaller when we see the blue graph line compared to the yellow line from 'Figure 3'. This indicates that it is better to approach this method than the previous three approaches.