Edith Flores
ML Homework 2: Regression-A Least Squares Approach
February 28th, 2019

Homework 1: Regression

Part 1.

After calculating the weighted matrix, W, using the training file we will determine the function, average loss, and curve/scatterplot for the validation file for polynomial degrees 1 to 6. The results are as follow:

Degree	Function *	Avg Loss*	Curve/ScatterPlot
1	$f(x)=105.00-50.40x$	17217.42	Table 1.1
2	$f(x)=-31.17-51.96x+16.32x^2$	4070.73	Table 1.2
3	$f(x)=-29.27-5.37x+16.22x^2-3.10x^3$	814.92	Table 1.3
4	$f(x)=3.81-5.05x+3.24x^2-3.10x^3+0.60x^4$	6.00	Table 1.4
5	$f(x)=3.81-5.05x+3.24x^2-3.10x^3+0.60x^4-1.82e-05x^5$	6.01	Table 1.5
6	$f(x)=3.76-5.05x+3.28x^2-3.10x^3+0.59x^4-1.09e-05x^5+0.00x^6$	5.99	Table 1.6

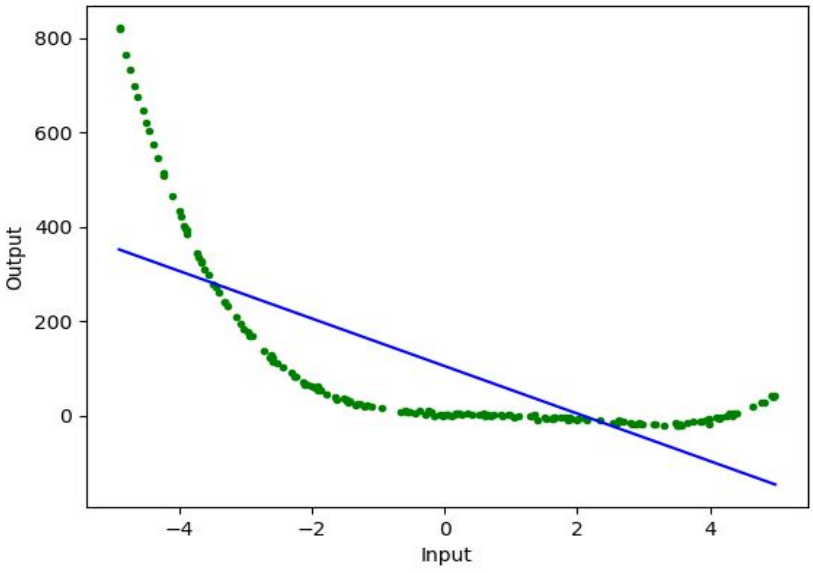


Table 1.1

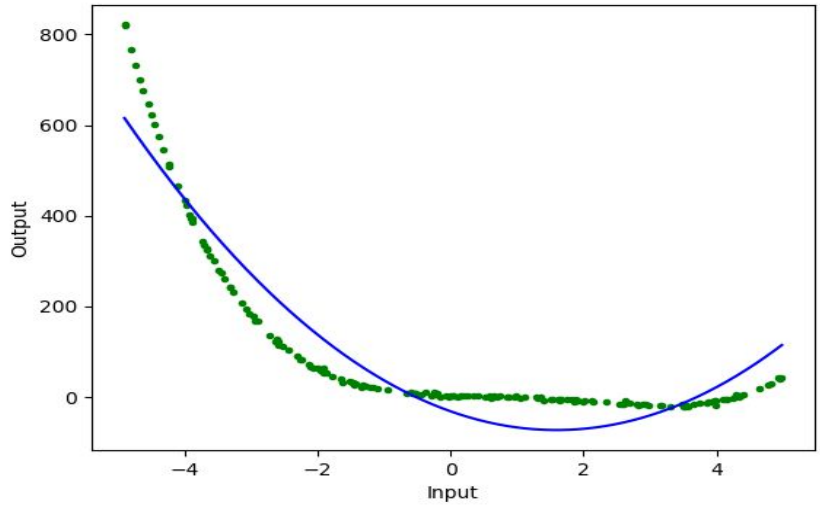


Table 1.2

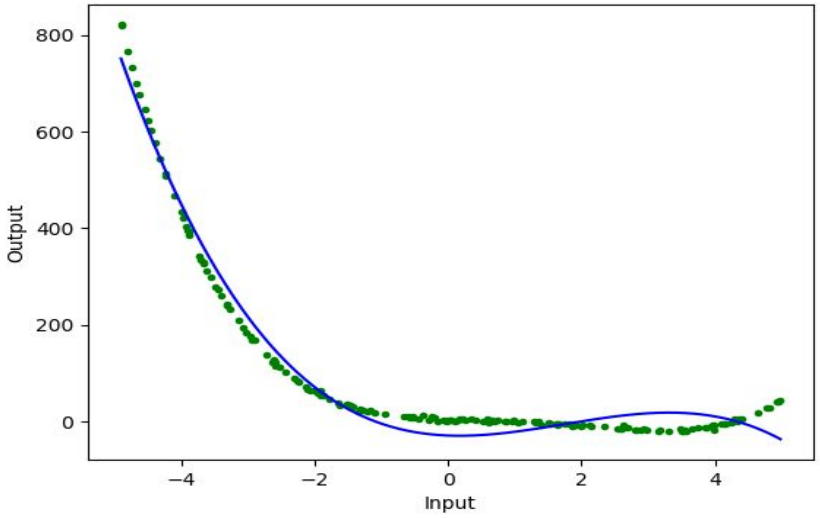


Table 1.3

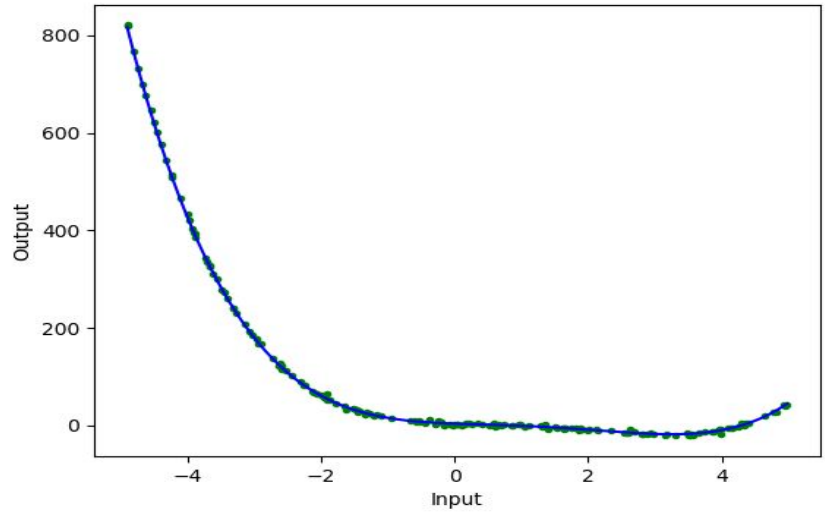


Table 1.4

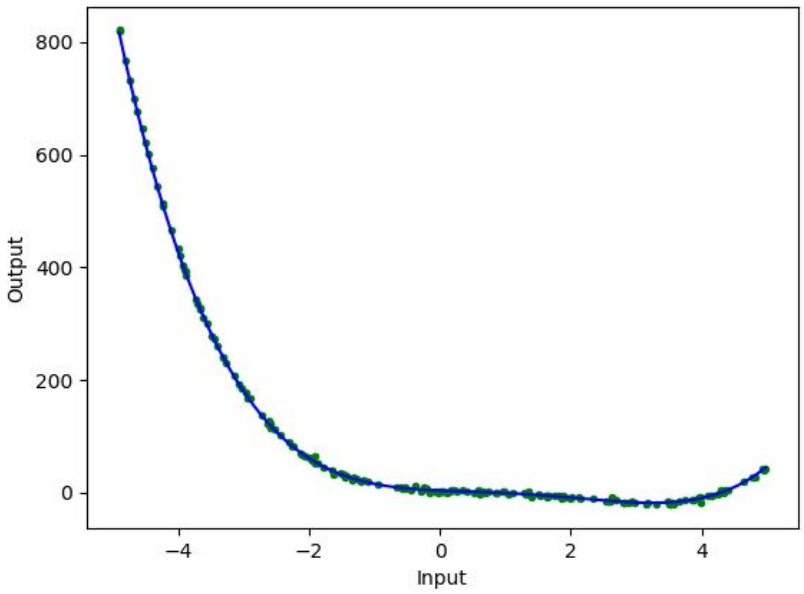


Table 1.5

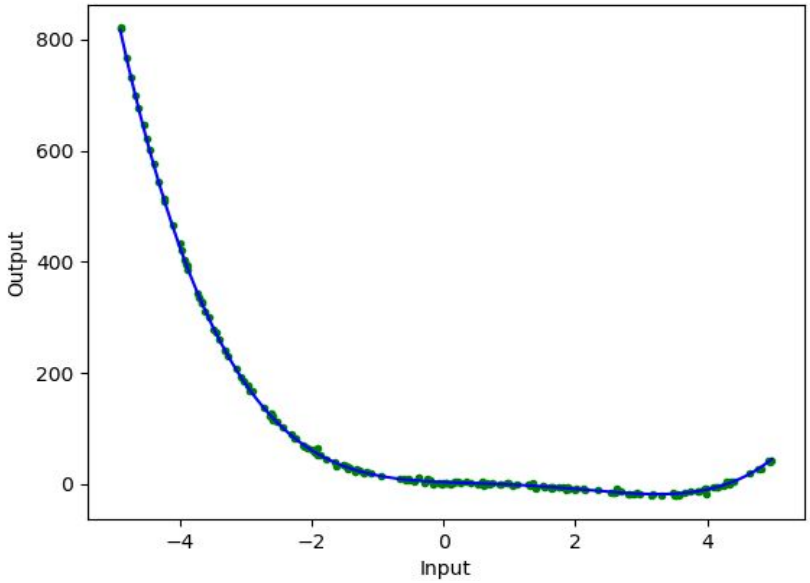


Table 1.6

Part 2.

Using the same method as Part 1, we ran the data file and did 5-fold cross-validation. After finding the average loss of each polynomial at each subsection, we calculated the overall average loss. The results are as follow:

Polynomial Degree	Overall Average Loss *
1	18912.88
2	4369.74
3	835.46

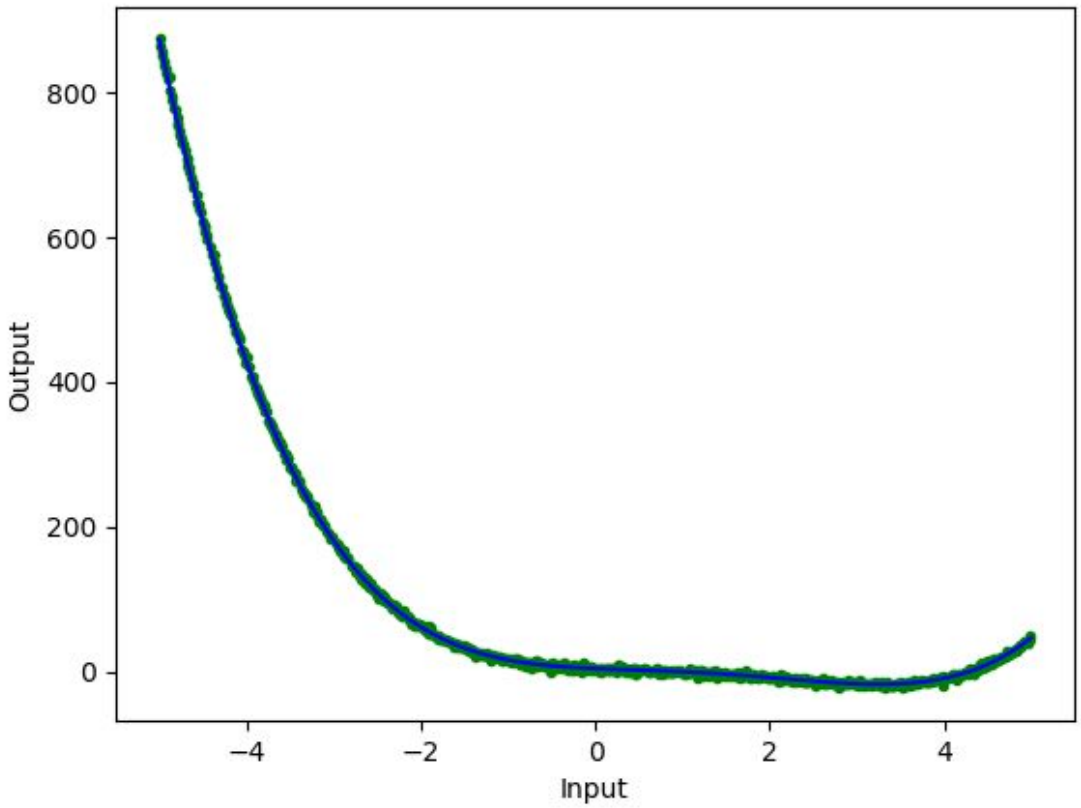
4	5.29
5	5.31
6	5.32
7	5.32
8	5.33
9	5.34
10	5.34

The lowest overall average loss is 5.29, which belongs to the polynomial of degree 4. Therefore, we would choose polynomial of degree 4 to be the polynomial we will use to do our predictions.

Part 3.

As mentioned under Part 2, polynomial of degree 4 is the degree with the lowest overall average loss. With this being said, we will use all the samples in the synthetic data file. Synthdata.txt, in order to find the final polynomial function of degree 4 and plot the curve with all the samples. The results are as follow:

* $F(x) = 3.89 - 5.04x + 3.23x^2 - 3.10x^3 + 0.60x^4$



*Function's constants and average losses are rounded to the hundredths place