Yonathan Mekonnen
Derrick Adeji
Nathan West
11/19/18

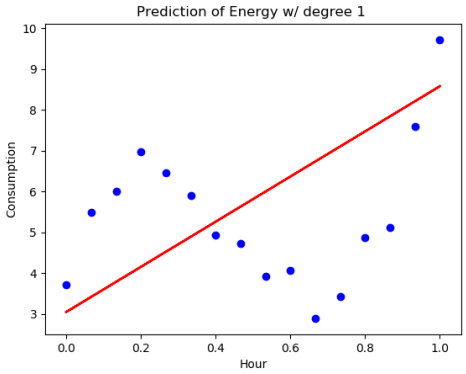
CMSC 409 Project 3 Report

1) The input for this problem is the hour intervals and the energy consumption in kW for that hour are the outputs.

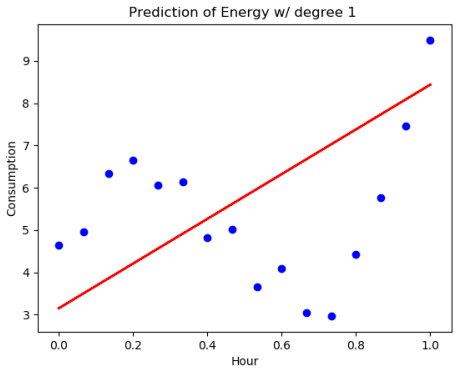
2) The activation function we chose is the delta rule with gradient descent. The delta rule is similar to the sigmoid function except the output is not limited to 0, 1, or -1. The data set we have does not have any classifications attached to it. Thus, using the delta rule made sense because our single neuron unit needed an output between -INF and +INF.

3) A) Polynomial Degree = 1

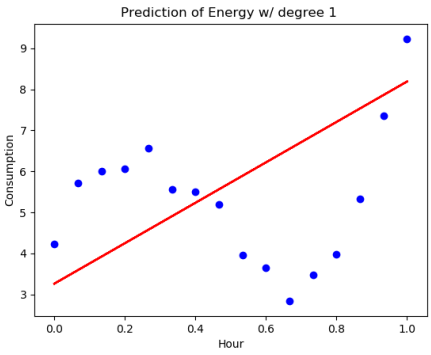
Day 1: Error = 44.85%



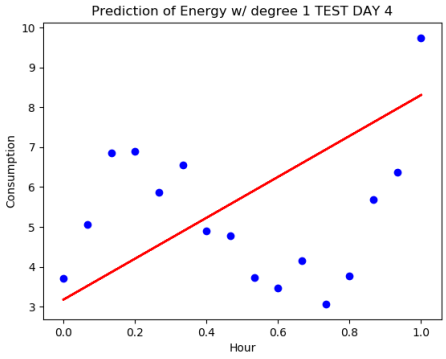
Day 2: Error = 42.48%



Day 3: Error = 39.19%

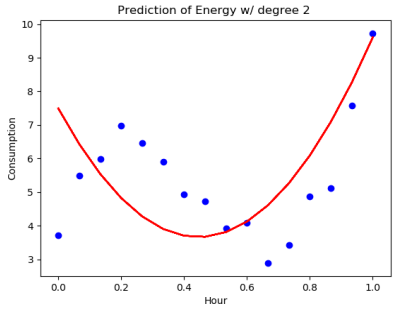


Day 4 (Test): Error = 48.14%

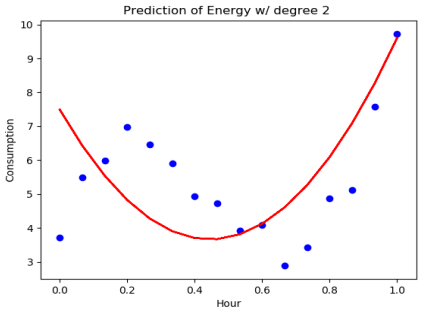


B) Polynomial Degree = 2

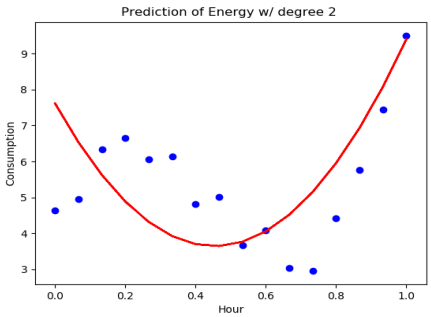
Day 1: Error = 40.21%



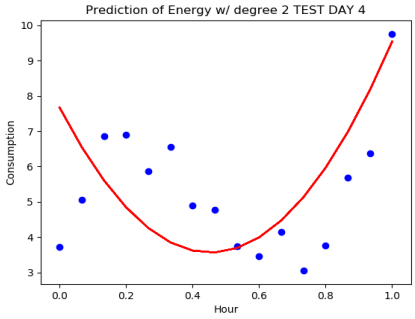
Day 2: Error = 33.16%



Day 3: Error = 32.58%

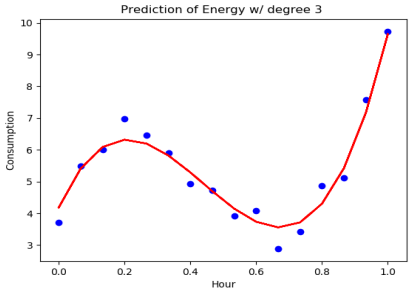


Day 4 (Test): Error = 47.89%

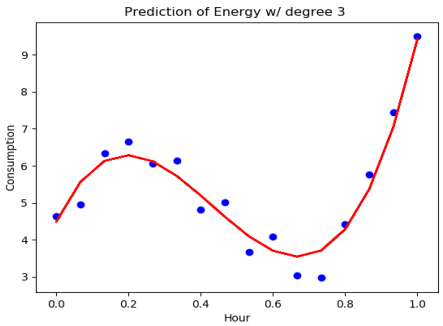


C) Polynomial Degree = 3

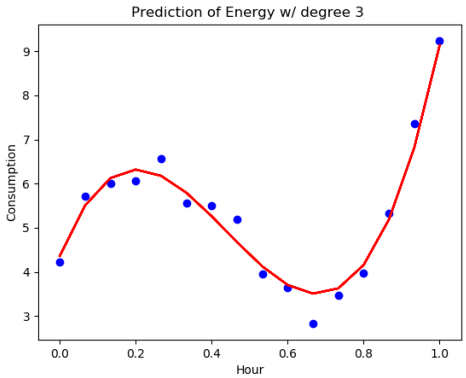
Day 1: Error = 3.65%



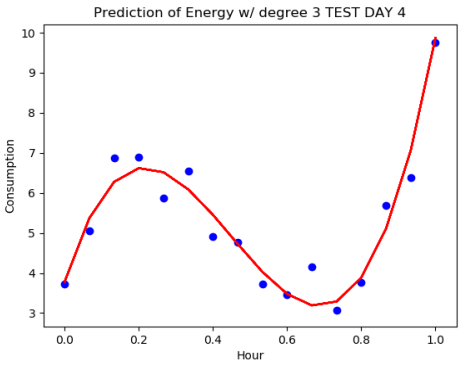
Day 2: Error = 3.48%



Day 3: Error = 2.05%



Day 4 (Test): Error = 5.96%



4) Number of iterations: 4000
Reasoning: With a small number of iterations, such as 500, the error rate of the single neuron unit greatly increases when the polynomial degree is 3. As the number of iterations increases, the error amount decreases and the graph sharpens to it's data points. We chose 4000 iterations because it seemed to be the lowest number of iterations that yields the lowest error amounts.

Learning rate: 0.3
Reasoning: We chose 0.3 as a learning rate because as the learning rate increases the error amounts are all over the place. If the learning rate is 0.1, the graphs when degree equals 3, are very different. Some draw the decision line perfectly along the original data, while others are drastically effected. The learning rate, 0.3, seemed to have a perfect balance and yielded more practically and trustworthy results.

Steps for Data Pre-processing: We normalized the hours so that we could compare numeric values obtained from different scales.

5) Yes, because the number of layers in neural network is greater than a perception thus the neural network will be able to learn much more out of the data thus can accurately predict given a new unseen instance