Jonathan Samson

Brandon Chin

CMSC 409 – Project 3

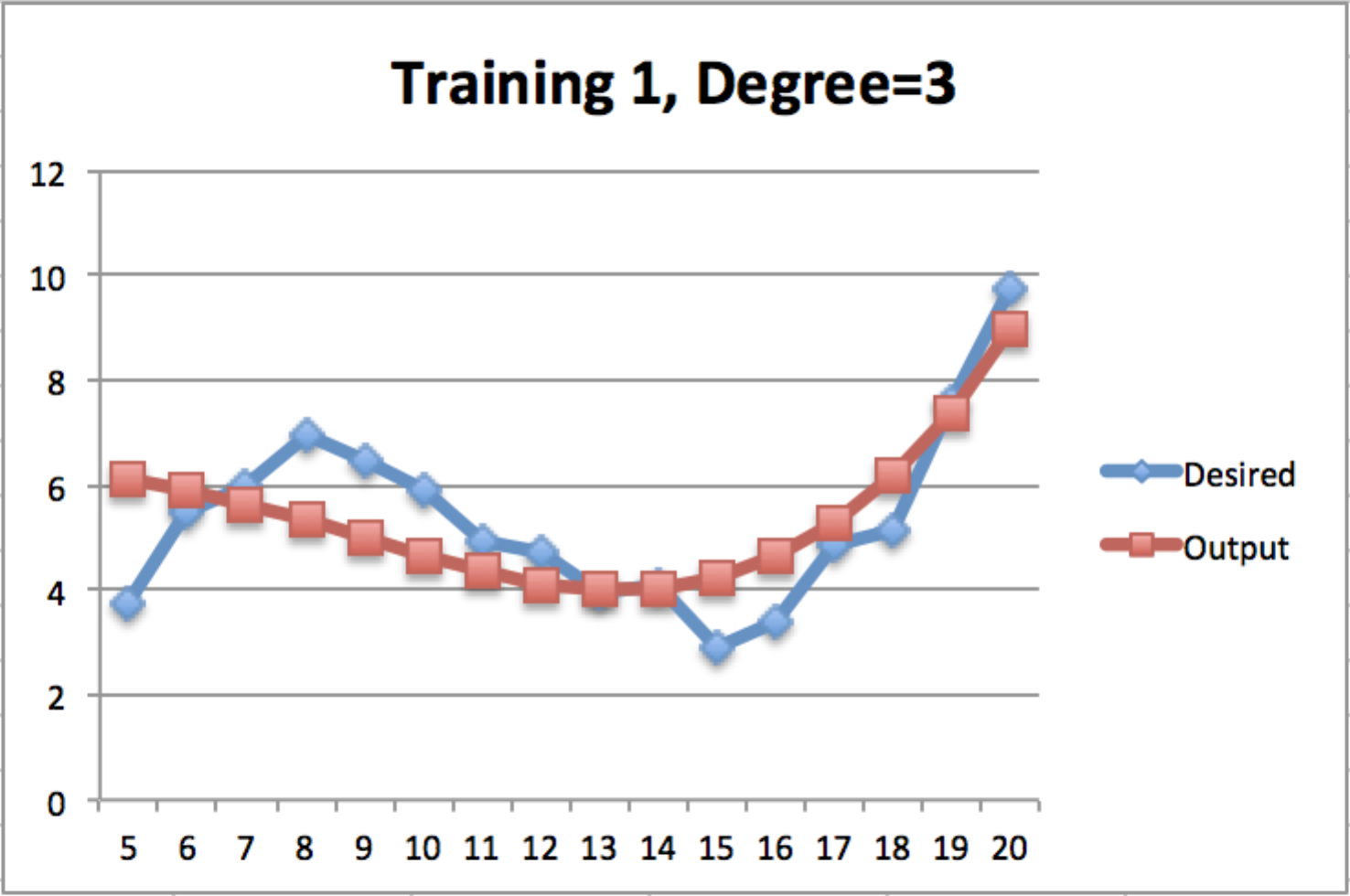
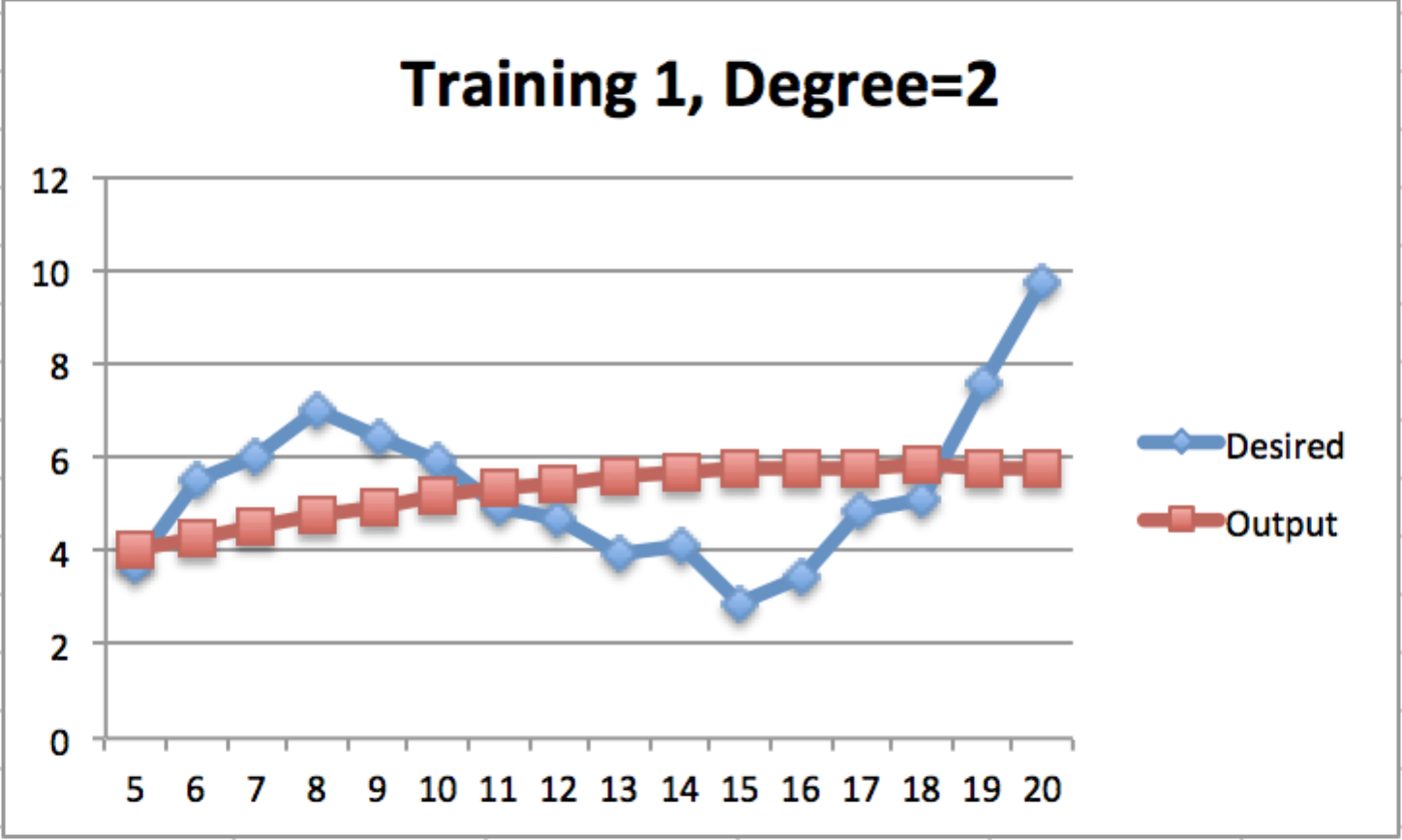
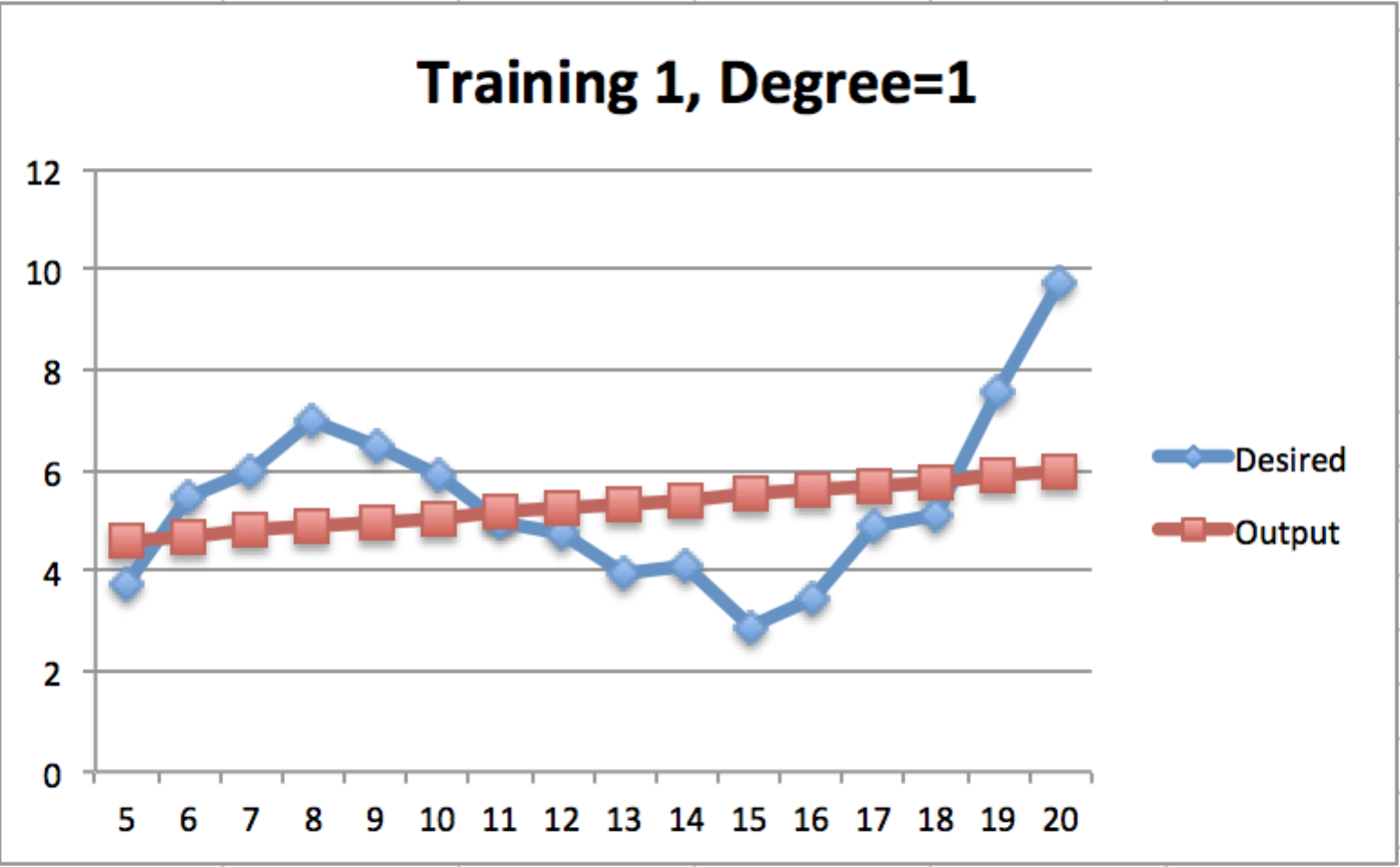
The inputs for this problem are the hours, and the outputs for this problem is the rate. The activation function of the decision unit is soft activation. This is because we are looking for continuous real values as outputs that are linear from negative infinity until positive infinity. However, these values should not be asymptotically like they would be in hard activation, and therefore soft activation makes the most sense for this problem. The number of iterations we decided to have is 1,000,000, and our learning rate we decided to be .00000001. We needed to make sure the training constant is low. After testing both the training constant and the number of iterations at smaller and larger numbers, we found that too few iterations or too large of a training constant, and the output would be wildly inaccurate and exceedingly large. Using our knowledge from the previous assignments as well as significant guessing and checking, we decided that the numbers we chose represented the most accurate output. We tried with only 100 iterations as well as with a learning rate of 0.5, and we saw that the learning rate needed to be much lower, and the iterations needed to be much higher in order to achieve a more precise output.

There are several data preprocessing steps that can be taken including: data cleaning, data integration, data transformation, data reduction, and data discretization. Data cleaning is filling in missing values, smooth noisy data, identify or remove outliers, and resolve inconsistencies. Given the data sets that were given to us, we did not have to do any data cleaning with the data. Data integration is using multiple databases, data cubes, or files. We are using data integration with pulling training data from multiple files as well as using a separate test data file. We do not have multiple databases or data cubes. Data transformation is normalization and aggregation. We considered normalizing the data, and for a few normalize. Though we tried to normalize the data, we felt that given the effects of normalizing the data to one and the resulting effects that it had on the bias, it was better to not normalize the data. Data reduction is reducing the volume but producing the same or similar analytical results. Data discretization is part of data reduction, but it is replacing numerical attributes with nominal ones.

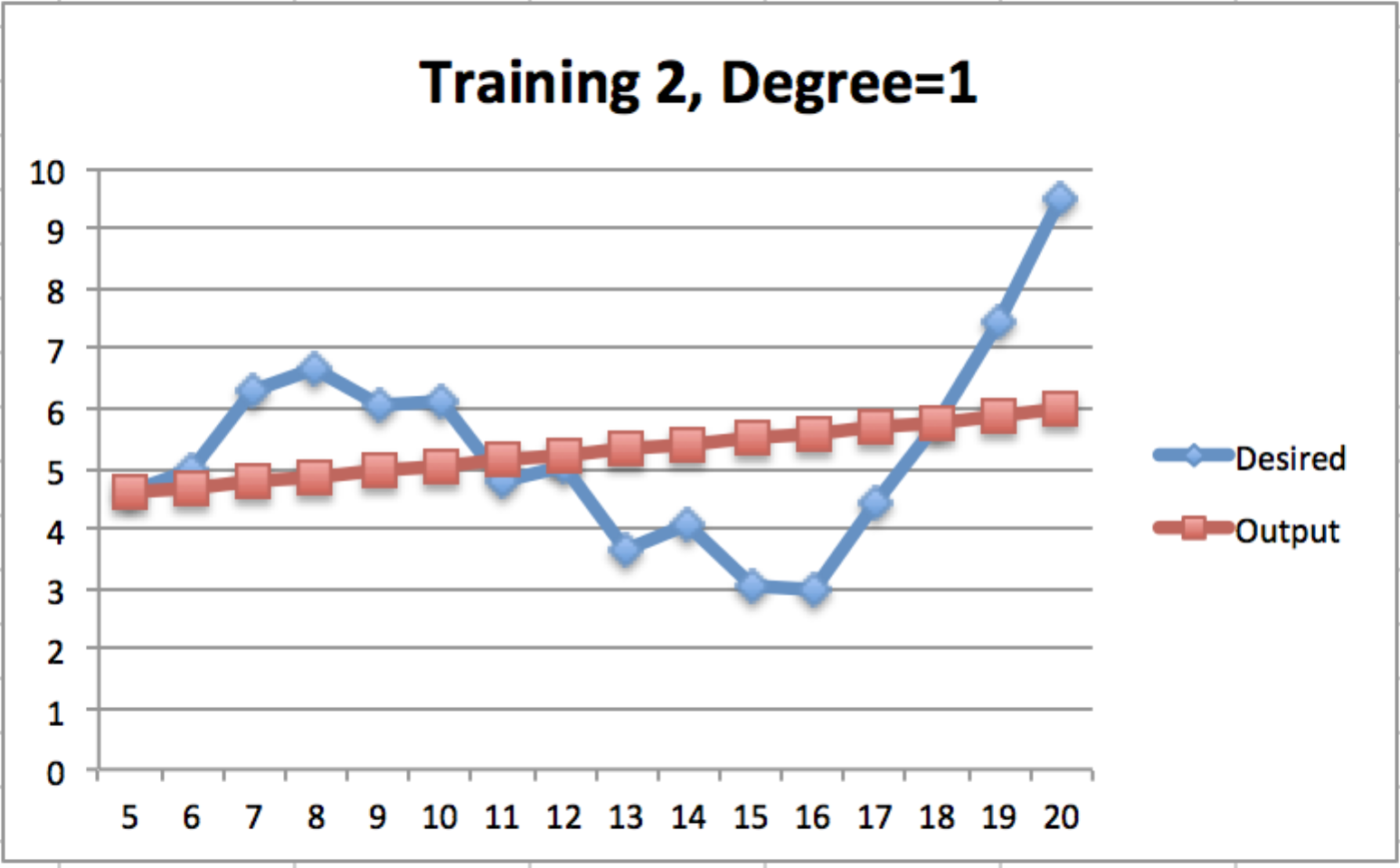
Could the error be further reduced using a network of neurons opposed to a single decision unit? If so, discuss how and why these methods would reduce the error. Yes, as the system gets more complicated, a network of neurons would be able to predict the trend more precisely. However, there would be more training required in order to reduce the error. This is due to an increase in the number of neurons and the need for each one to make a decision, which would mean that each one would need to be trained on the data set, and you would need significantly more data in order to efficiently train the network of neurons.

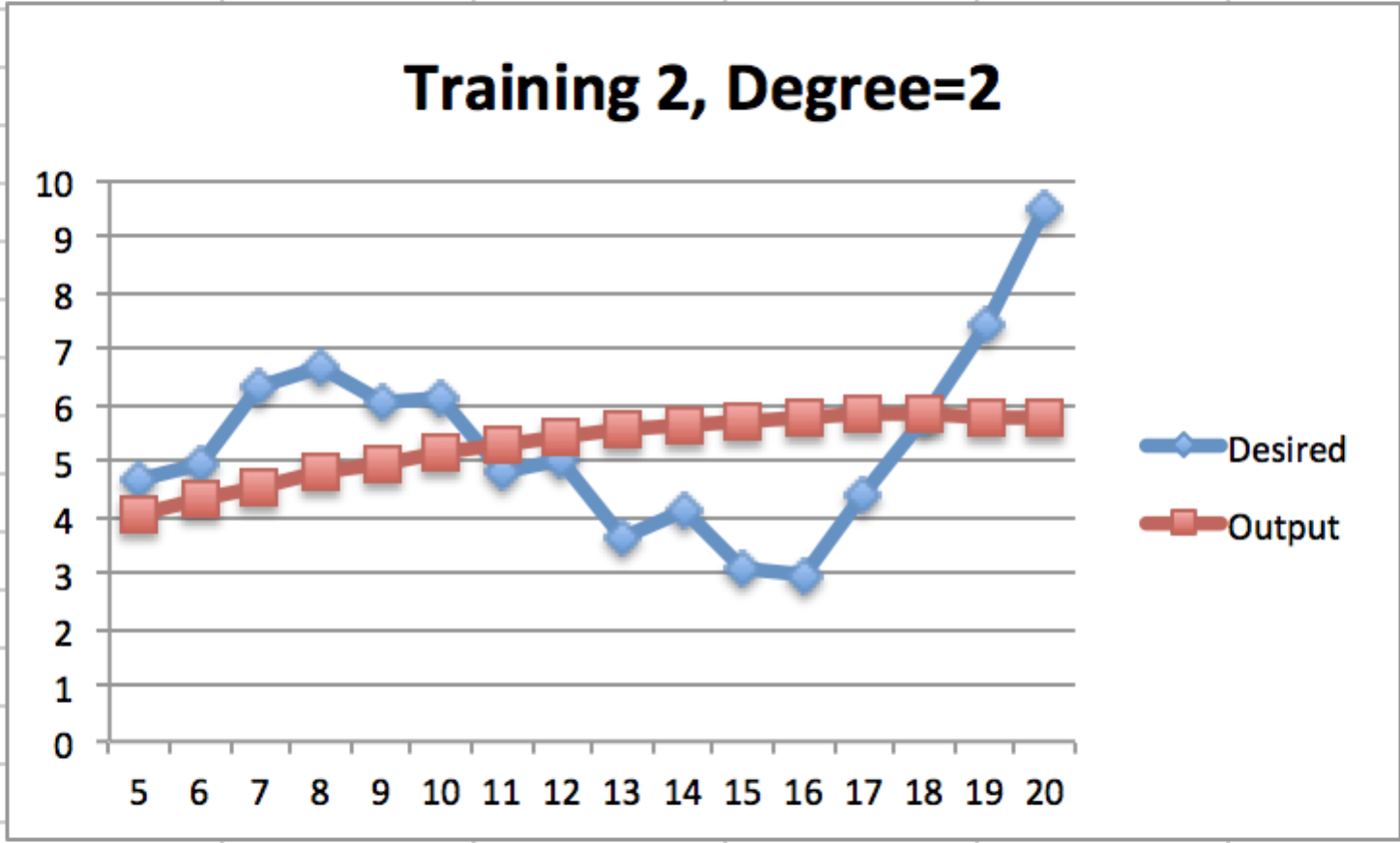
**Graphs**

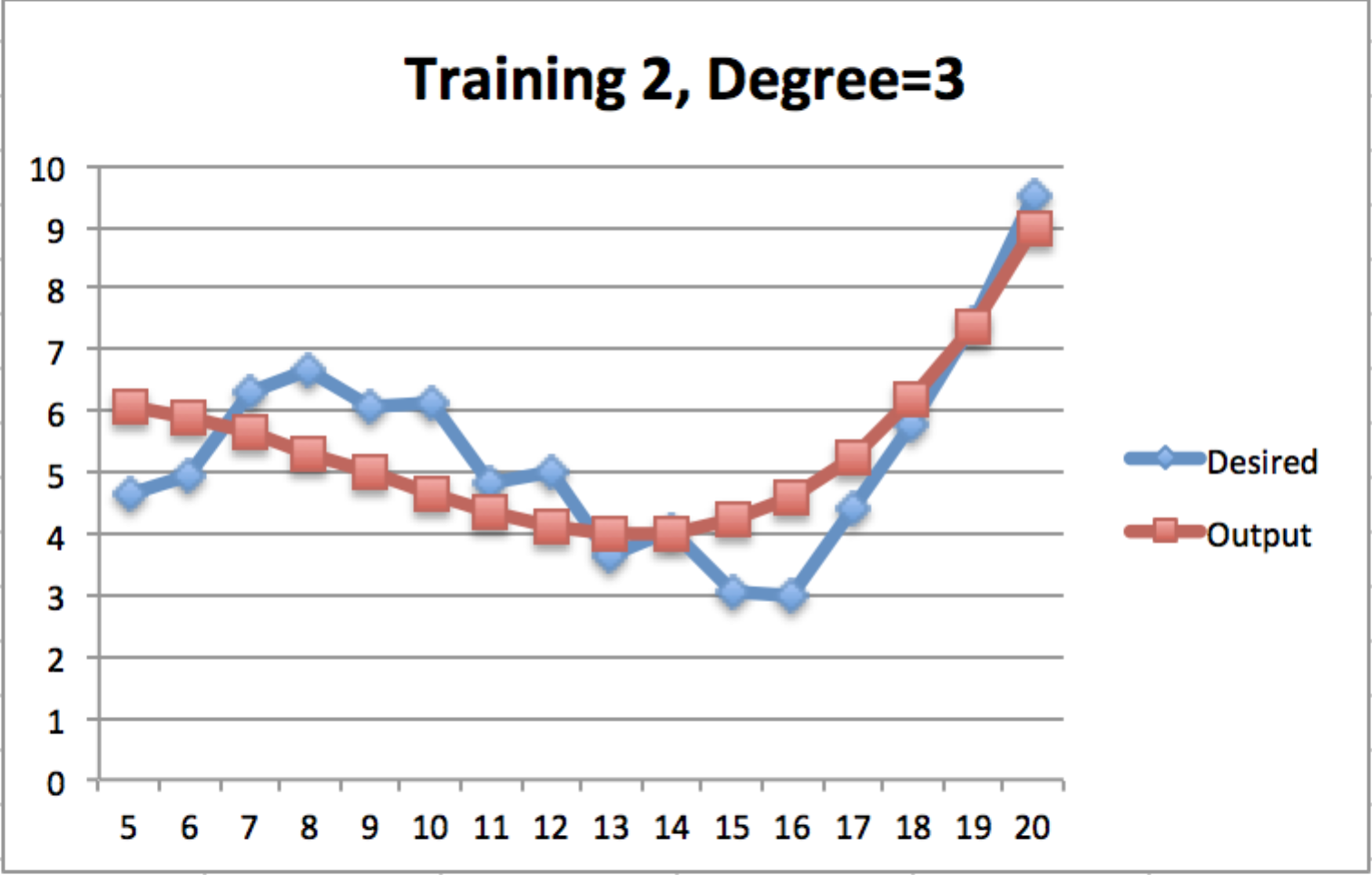
Day 1 (Training 1)



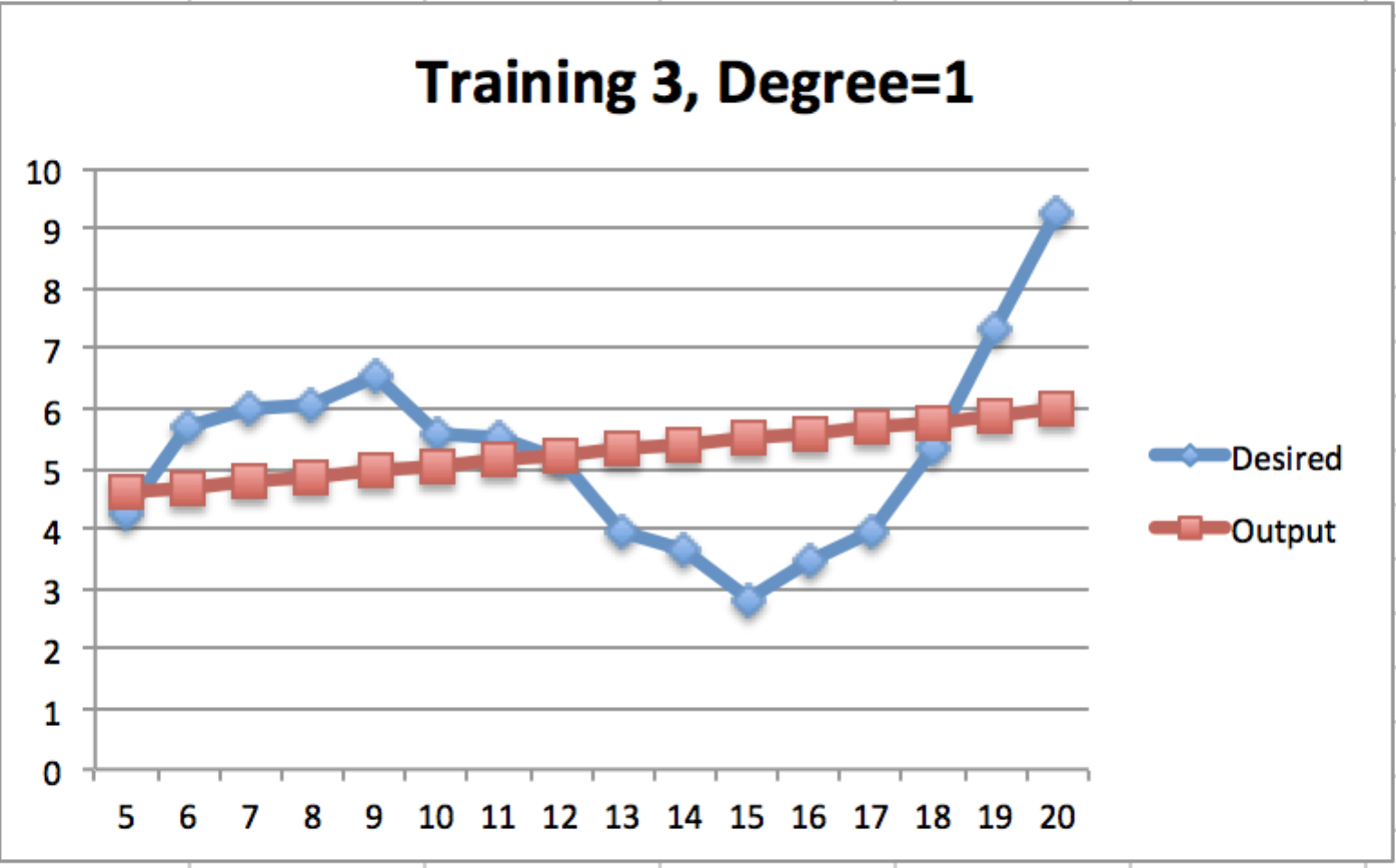
Day 2 (Training 2)

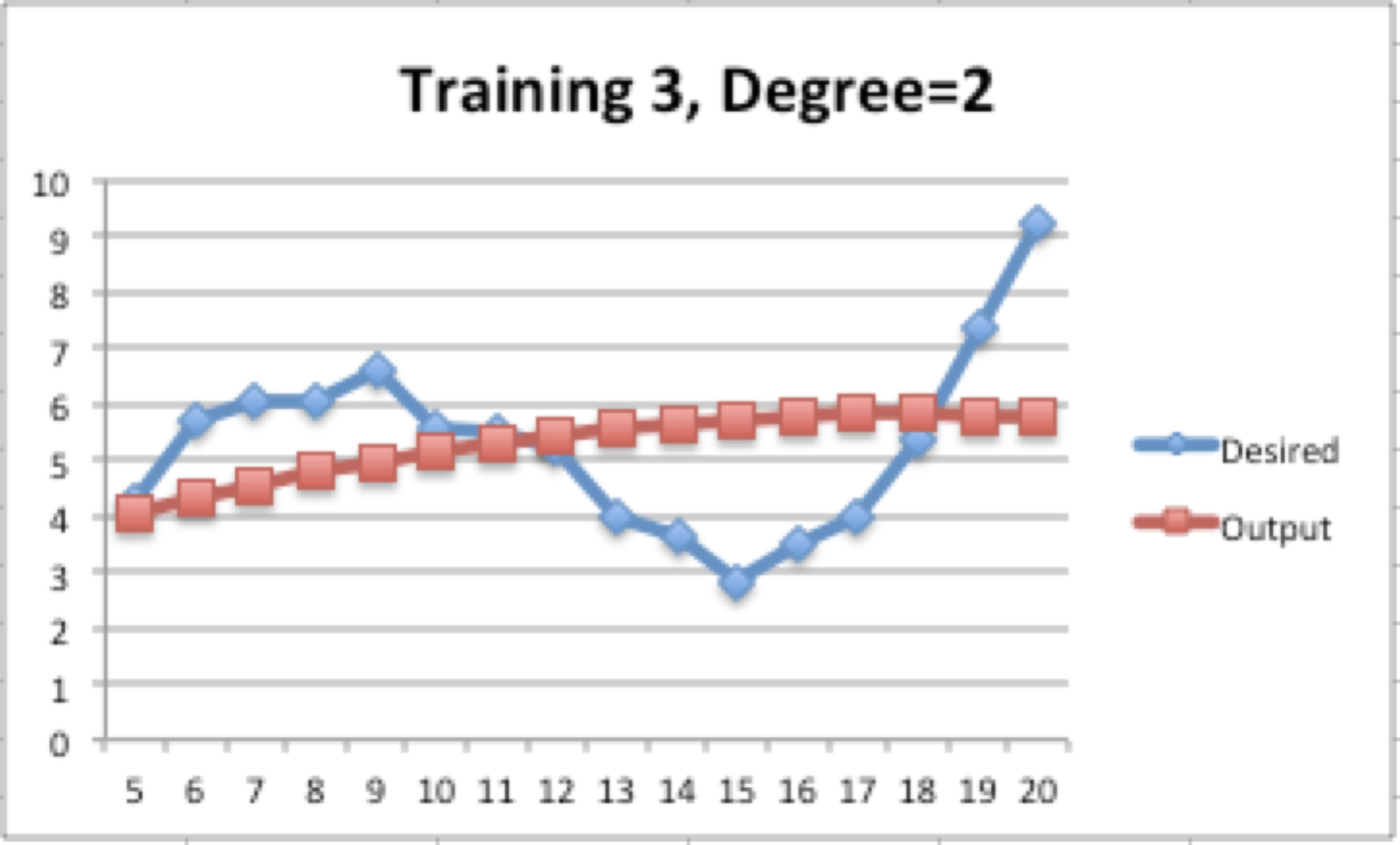


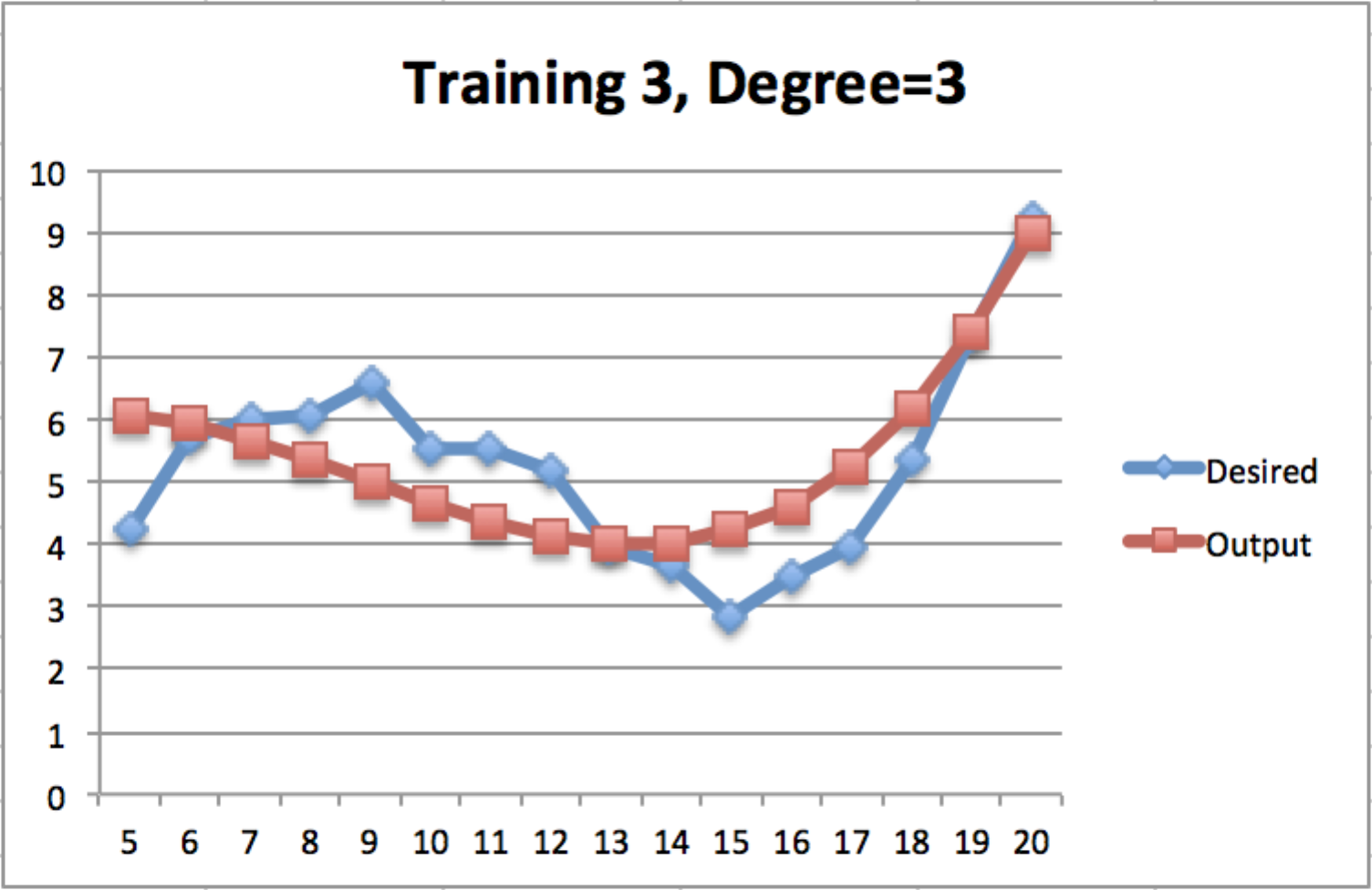




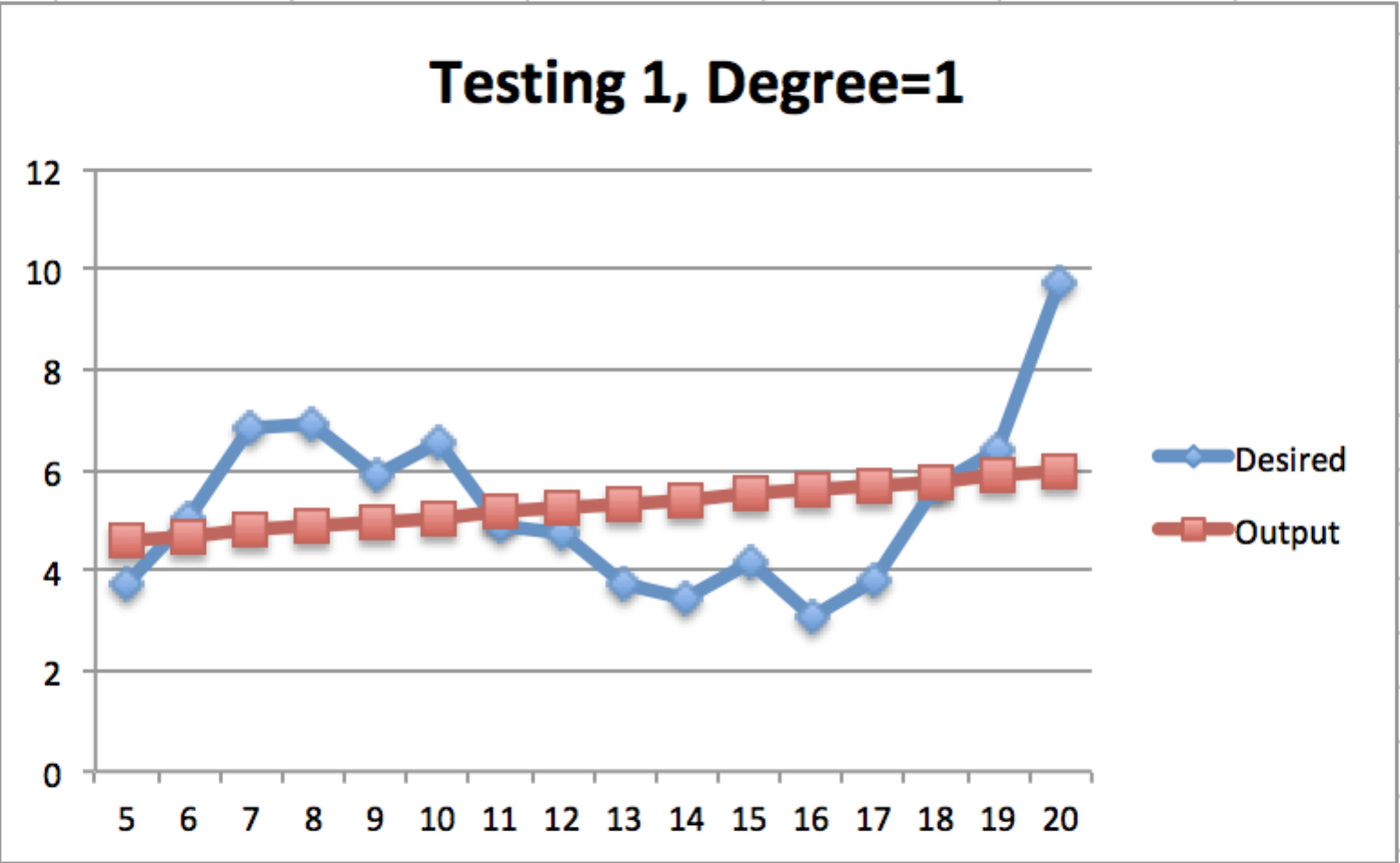
Day 3 (Training 3)

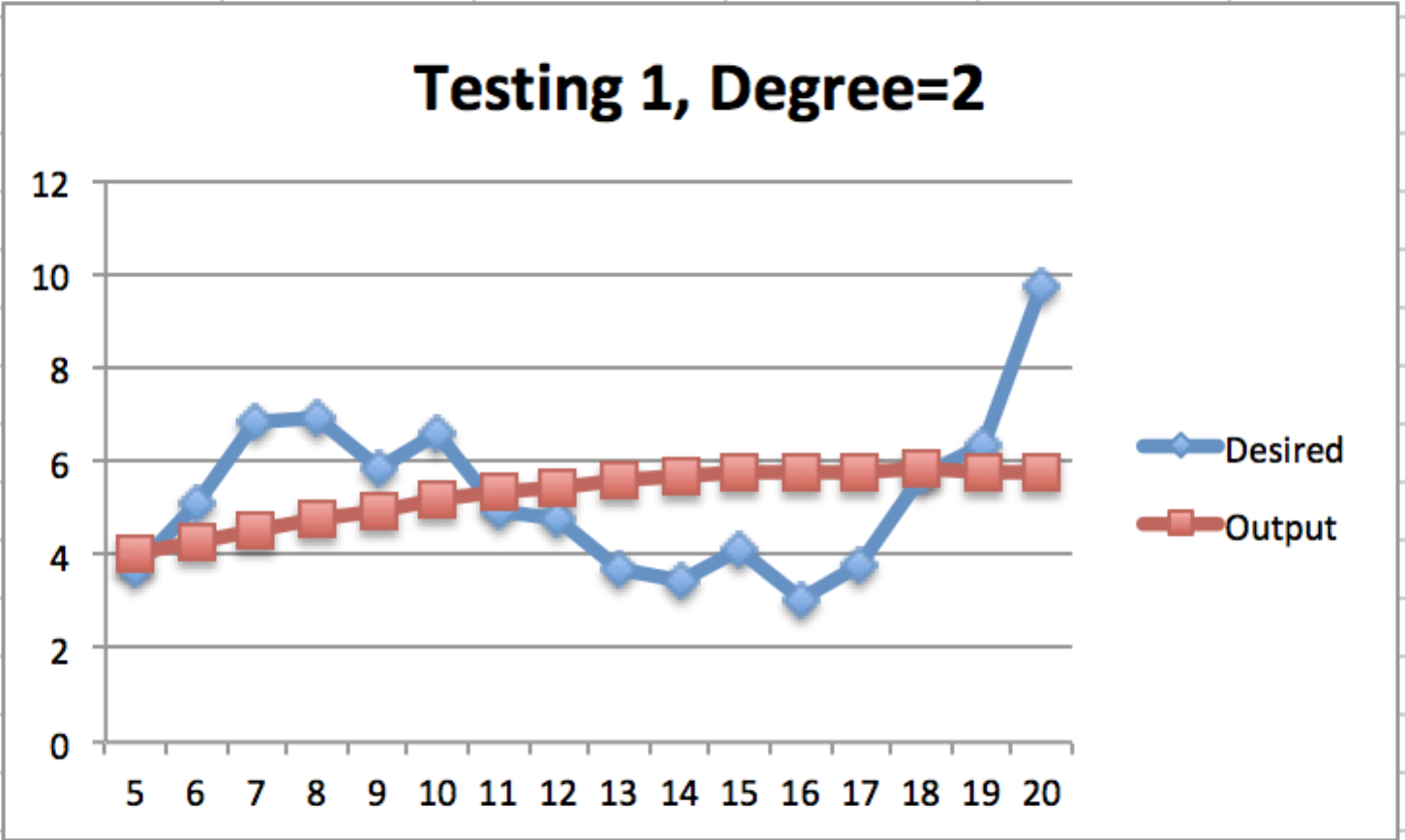


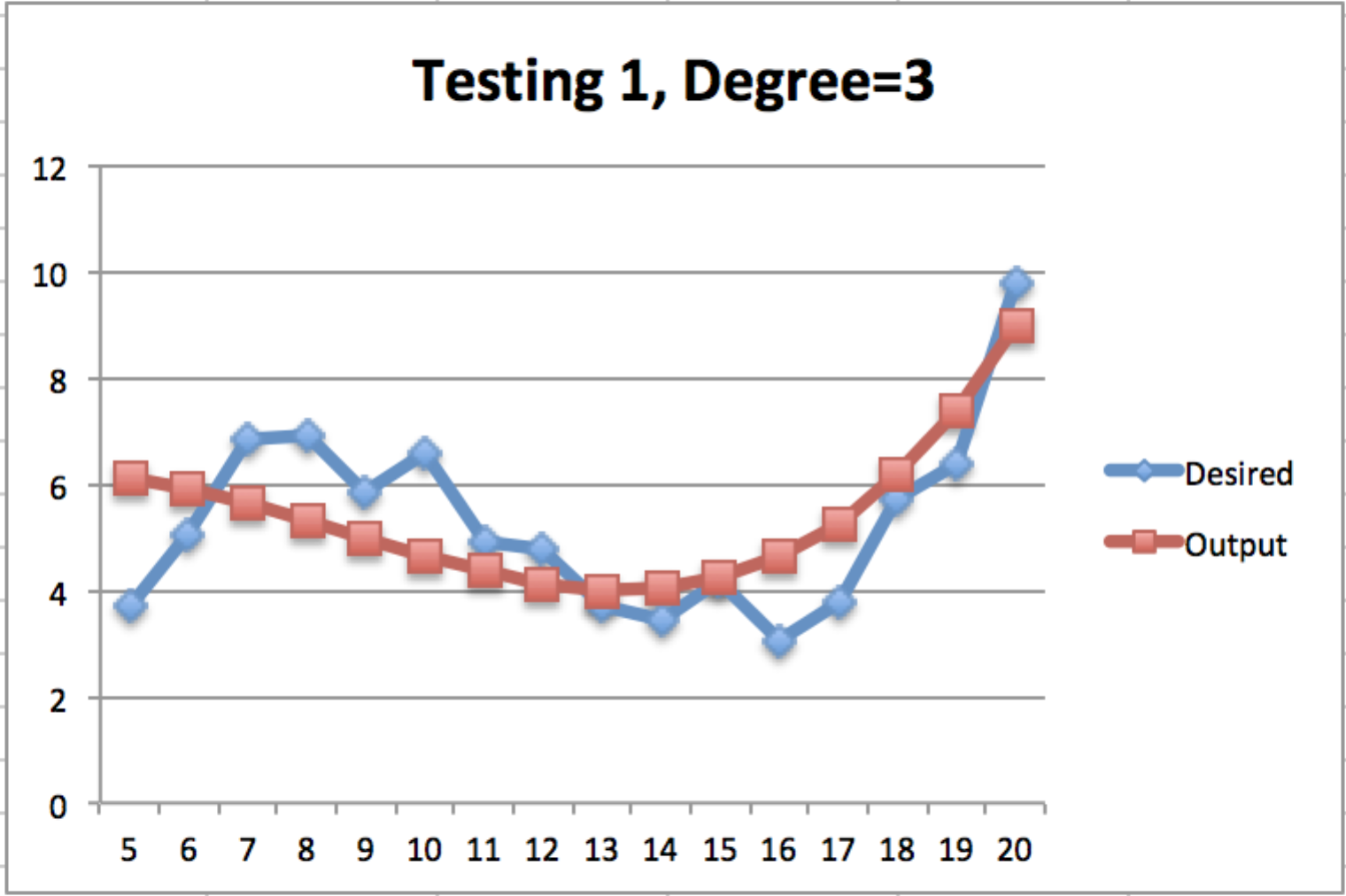




Day 4 (Testing 1)







**Total Error for Degree 1:** 11.2296014

**Total Error for Day 1:** 6.649214787984489

**Total Error for Day 2:** 6.5122696869325445

**Total Error for Day 3:** 6.318587571614721

**Total Error for Degree 2:** 12.15826526

**Total Error for Day 1:** 7.3437976265969676

**Total Error for Day 2:** 7.283992162956873

**Total Error for Day 3:** 7.068096708100044

**Total Error for Degree 3:** 6.99409024

**Total Error for Day 1:** 4.1492729726042255

**Total Error for Day 2:** 3.758825869971608

**Total Error for Day 3:** 3.849389354715179