

Homework 4

Morgan Baker

11/1/16

1 Question 1

This was the only question on the assignment. The question was to use the k nearest neighbors algorithm to find the best Eout with a 10-fold Cross Validation. The dataset generator didn't work with the nearest neighbors algorithm because the data needed reshaping. The algorithm was finished and i got an array of Eouts that were only organized by their index, or the number of nearest neighbors/2 - 1. The output to this array can be seen in Figure 1. The remaining task was to create an algorithm to find the best eout, or the smallest number, and get the number of nearest neighbors that produced the value. I realized that some of the best k-values were produced from the first few iterations, which surprised me. With some of the other points, however, the algorithm was definitely overfitting the data. The three best values, along with their n nearest neighbors, can be found in Figure 2. I ran this program a couple of times, and each time got similar results. There would be one early index that did really well, and one that did really well really late, but the third was usually a toss up. This homework was interesting because seeing this in action helped me learn something useful for my semester project, with finding different parts of numbers that can be recognized.

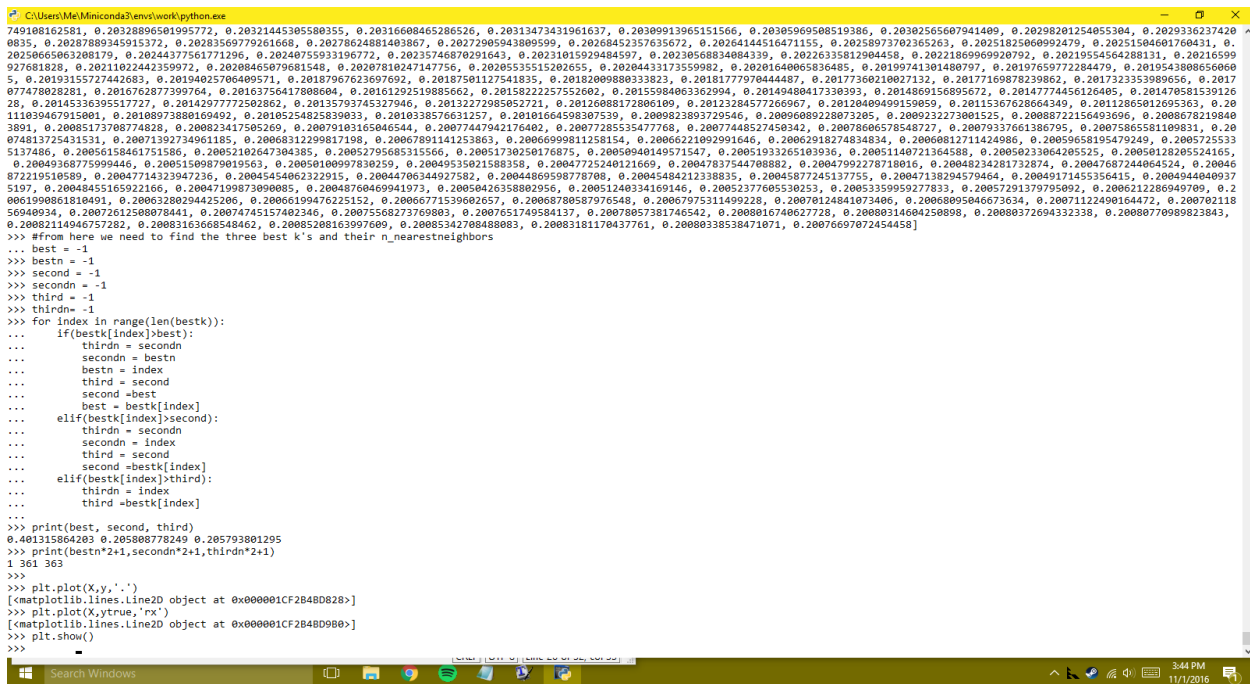


Figure 2: The three best values of E_{out} , along with their number of nearest neighbors.