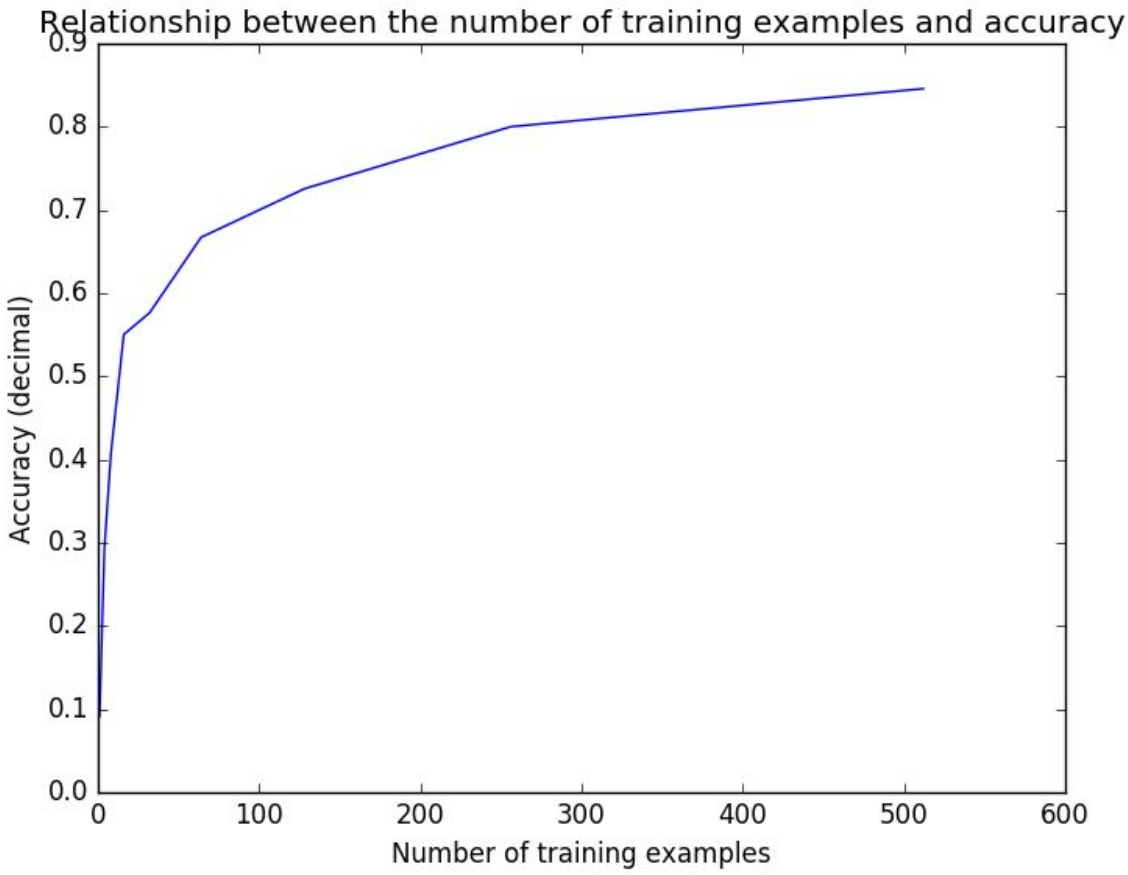
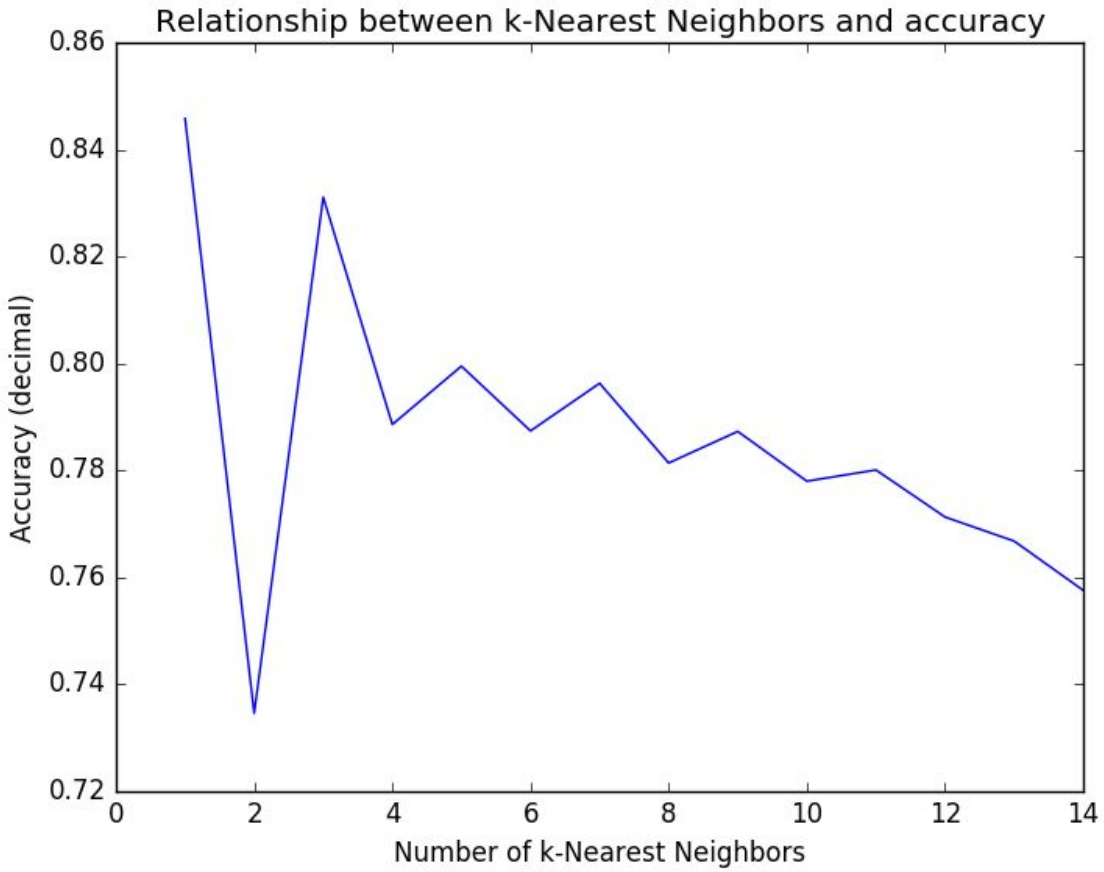
Robert Werthman
CSCI 5622
Homework 1: KNN

What is the relationship between the number of training examples and accuracy?
From the graph below, as the number of the training examples increases the accuracy of the classifier increases logarithmically. This means having more training examples gives you better accuracy with the KNN classifier, but at some point you get very little benefit from adding more training examples .



What is the relationship between k and accuracy?
From the graph below, as k, the number of Nearest Neighbors, increases the accuracy of the classifier decreases linearly. Based on the graph, you should use a smaller k to get more accurate labeling from the KNN classifier. The below graph used 500 training samples for each value of k.



What numbers get confused with each other most easily?
Classifying with only 500 samples from the training data, the numbers that get confused the most with other numbers are:

- 0 confused with 6, happened 26 times
- 1 confused with 5, happened 5 times
- 3 confused with 5, happened 63 times
- 4 confused with 9, happened 155 times
- 5 confused with 3, happened 82 times
- 6 confused with 1, happened 34 times
- 7 confused with 9, happened 95 times
- 8 confused with 3, happened 63 times
- 9 confused with 4, happened 60 times

The ones that get confused the most usually look very similar when written down, contain a lot of the same curves, and have around the same number of line segments. There could also be mislabeling due to the fact that we take the median of the majority labels if there are multiple majority labels. This could end up labelling numbers incorrectly.