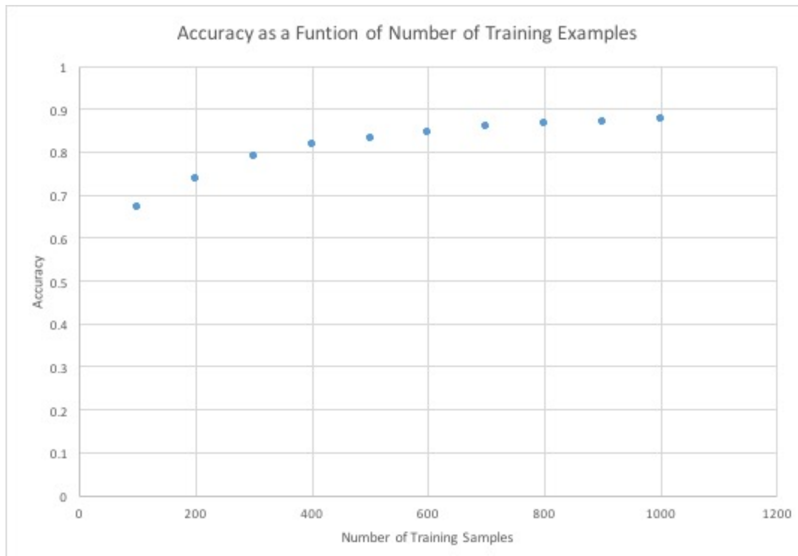
Donovan Guelde
KNN HW
CSCI 5622

1. What is the relationship between the number of training examples and accuracy?

Intuitively, as the number of training examples increases, so does accuracy. This increase happens relatively quickly while training sample numbers are in the low hundreds, but levels off as accuracy begins to approach .9. With no limit on training samples and k = 3, accuracy = .9712, so the pattern appears to hold as the limit approaches max.

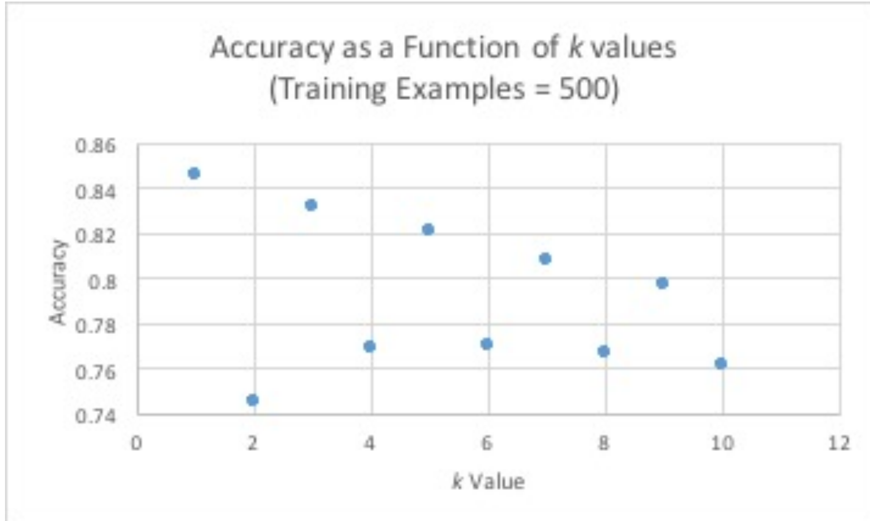
# training examples(with k=3)	Accuracy
100	0.67
200	0.7382
300	0.789
400	0.8177
500	0.8311
600	0.8436
700	0.858
800	0.866
900	0.8685
1000	0.8758



1. What is the relationship between *k* and accuracy?

With a constant number of training samples (500), accuracy generally decreased with increasing *k* values, with odd *k* values (1,3,5…) producing noticeably better accuracy than even values (2,4,6…). Accuracy was increased with more training samples, but the same relationship between even and odd k values appears to hold.

K Value (training samples = 500)	Accuracy
1	.8458
2	.7455
3	.8311
4	.7686
5	.8203
6	.7694
7	.8081
8	.7665
9	.7966
10	.7616



2. What numbers get confused with each other most easily?

The most commonly confused numbers appear to be 9 and 4. This confusion appears to persist across all *k* values and number of training examples. As an example, where *k* = 3 and training examples are limited to 500, 4 is improperly identified as 9 155 times, and 9 is misidentified as 4 60 times. These two error categories alone account for 12.73% of the total error (215 of 1689 errors total). This proportion of total errors for the 9/4 combination appears to be consistent. With k=3 and no limit on training samples, the 9/4 combination gives a total of 163 out of 1263 total errors, or 12.9% of all errors.