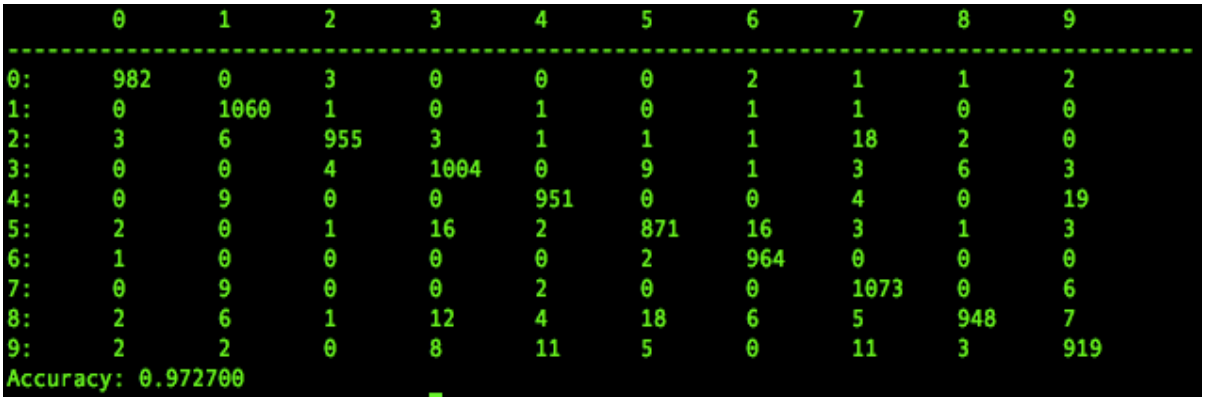
KNN Classifier

Anas Salamah
810-87-6641

1. What is the role of the number of training points to accuracy?
   - It seems that the higher the number of points the better the accuracy. Limiting the data points to 500 gives me an accuracy of 0.831100. At 10 points the accuracy decreases dramatically which makes sense since the classifier needs more data to approximate the labels more accurately. The good thing is that when using all the data points, the accuracy is at 97% that proves that with more data, the KNN classifier is a very confident one.
2. What is the role of ~k~ to accuracy?
   - There is no rule to what the k should be for optimal accuracy. k is data dependent and there is no correct k for the best accuracy. I found that an even k is usually less accurate than an odd k that is close to the even number. I also found that when running the whole data set, k=1 produced accuracy that is a little less than the default k. However, on 500 points k=1 was doing better than k=3. From my analysis, it looks like when you have less data points, k=1 has the most accuracy, but with an ample of data, an odd k that is slightly bigger than 1 is better.
3. What numbers get confused with each other most easily?
   - When using the whole data set, when the true label is "9", the algorithm labels it incorrectly as "4" 19 times. Two close 2nd most easily confused numbers are "5" as "8", and "7" as "2". Here is what it looks like:



	0	1	2	3	4	5	6	7	8	9
0:	982	0	3	0	0	0	2	1	1	2
1:	0	1060	1	0	1	0	1	1	0	0
2:	3	6	955	3	1	1	1	18	2	0
3:	0	0	4	1004	0	9	1	3	6	3
4:	0	9	0	0	951	0	0	4	0	19
5:	2	0	1	16	2	871	16	3	1	3
6:	1	0	0	0	0	2	964	0	0	0
7:	0	9	0	0	2	0	0	1073	0	6
8:	2	6	1	12	4	18	6	5	948	7
9:	2	2	0	8	11	5	0	11	3	919

Accuracy: 0.972700