3. When we split the numpy array into two parts with the training set being the top 80% of the data and the test set the last 20%, the accuracy was 96.88%.

4. When we swap so that the first 20% is the test set and and the training set is the remaining 80%, the accuracy drops to 91.37%

5. yes it makes sense that the first 5 have been missclassified because they are very unclear.

6. How the data was collected:

link:https://ocw.mit.edu/courses/15-097-prediction-machine-learning-and-statistics-spring-2012/d1cfd95258db2d252fd921b39805907d_digits_info.txt

the sample was was collected from 250 samples by 44 writers. A WACOM PL-100V pressure sensitive tablet was used with an integrated LCD display and a cordless stylus. The tablet sends x and y tablet coordinates and pressure level values of the pen at fixed time intervals (sampling rate) of 100 milliseconds. Some potential issues: since there is only small amount of writers, the data might not represent the variability of writing styles. The small writer-independent test set may not adequately evaluate generalization. The model may not generalize well to other hardware setups.

8. I chose k = 20 arbitrarily and the resulting accuracy was 88.97%

9. K is not always the same for all of the random seeds. In all 3 of the random seeds we chose though, the ks with the best results were k values 1-3.

10. The accuracy using compareLabels is 94.34% with a best_K being 3.