Assignment: Multiple features for digit classification

February 23, 2011

This assignment extends the previous assignment in which you classified 3's and 5's.

You will use the same training and testing data. The difference is that you will be using multiple pixel features instead of a single pixel feature.

1. First you will extend your classifier so that you can handle multi-pixel features. Consider the case of two-pixel features. You will select these pixel locations by hand, as you did in the last assignment. Call the first pixel A and the second pixel B. With the training data, for each class (3's and 5's), you will need to measure the frequency of all possibilities for the two features. In other words, you will need to measure $P(A = 0.5, B = 0.5), P(A = 0.5, B = 255.5), P(A = 255.5, B = 0.5), P(A = 255.5, B = 255.5)$. This is just a table of four different frequencies which should add to 1.

2. Once you have such a table for each class (these are the likelihoods, you should be able to use your classifier that you wrote in the last assignment. Note that you can turn your 2x2 table into a vector of four numbers by using the matlab colon operator. This should make it easier to use the code for your old likelihoods, which were vectors rather than tables.

3. Try experimenting with different numbers of features. Each time you add a new feature, use the same set of features you used before, but add one new feature. That way, you can study the benefit (or harm) of adding another feature without worrying that you might have picked inferior features.

4. Report your results for up to 10 features. How does performance change as you add more features? Make a plot of performance versus number of features using the plot function in matlab.

5. Also, make a figure that shows the features you used with a number at approximately each feature location. Do this using the matlab "text" command. In other words, put a "1" in the location of the first feature, a "2" in the location of the second feature, and so on. It would be good to do this on an image of the average 3's or the average 5's, or even on

the image of the difference between the averages, so I get some feeling for what the features mean.

6. Turn in your code, your plots, your images, and a short discussion of results to Manju.