Assignment 3: Visualize using PCA.

The data dat_2_3.csv contains 214 handwritten 2's and 3's images scanned from postal envelopes, like the ones shown below.

These images are stored as a 214X784 matrix. Each row of the matrix is an 28X28 greyscale image. Figure 2 (below) illustrates the two most significant dimensions found by PCA.
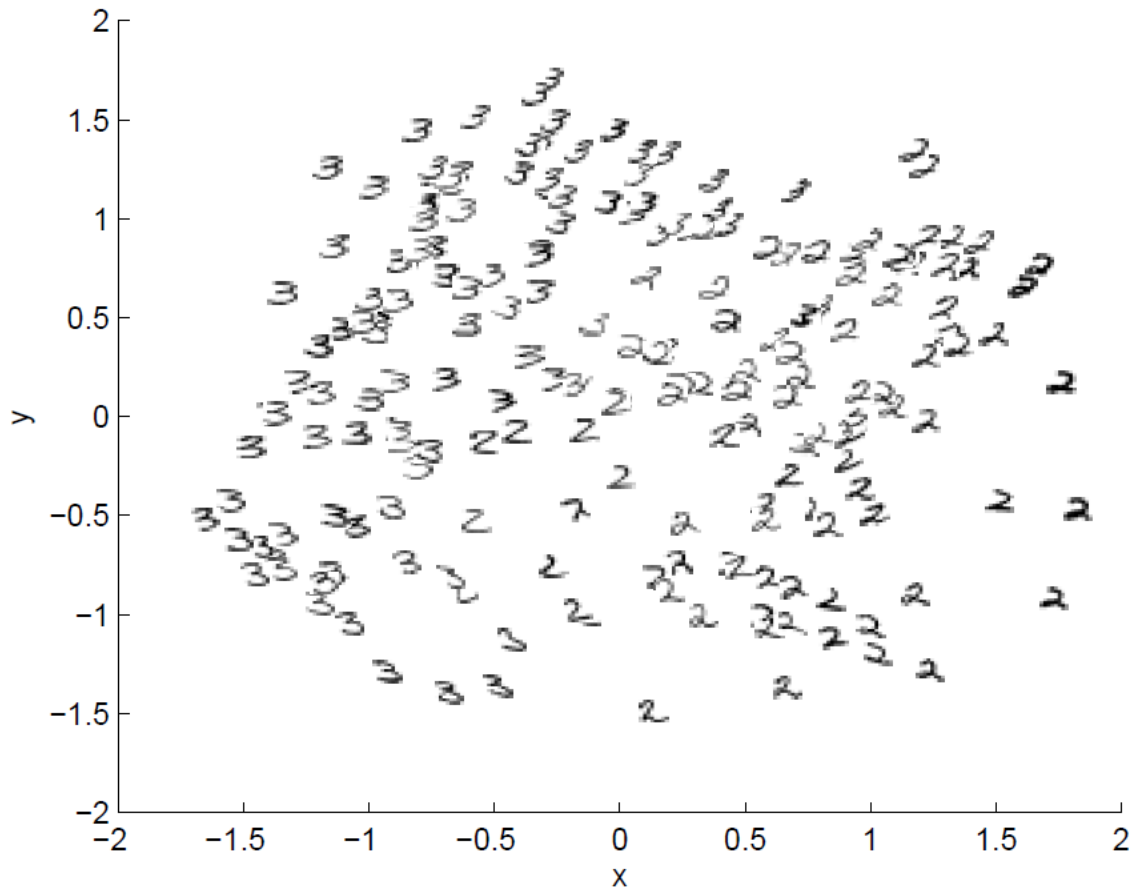

Figure 2

a. Reproduce Figure 2. You only need to find the low-dimensional mapping by PCA, and then plot the digits. You may represent them using shape or color.

b. Plot all combinations of the first five dimensions (for example, the first dimension vs. the third one, or the second one vs. the fourth one) and find the dimensions which separate 2's and 3's the best. In other words, Plot 10 plots for all 10 possible combinations. Then identify the one which best separates 2's and 3's and don't forget to label axes.

c. Repeat b but using Kernel-PCA with an appropriate kernel (and parameters) from your own choice. If you choose the kernel and its parameter wisely you should observe improvement over PCA.

d. As you probably observed, the low-dimensional map was affected by the choice of kernel function. Why does the kernel that you have chosen work better than the others?