M676 Pattern Analysis

Problem Set Six

Thursday, May 13, 2004

Cats and Dogs

This assignment will be collected during our special class on Thursday, May 13 at 1pm (room t.b.a.). Note that this is not the usual finals time so please confirm you have no conflicts while there is still time to change our meeting time.

To begin, load classdata.mat from the webpage. This contains two variables: the 198×198 matrix KLDATA and row vector sub-labels of length 160. The data matrix KLDATA has images of cats and dogs (courtesy of Ross Beveridge, Department of Computer Science). There are 99 of each animal and they are randomly placed in the columns of KLDATA. The vector sub-labels gives you the identity (i.e., cat=1 or dog=0) of the first 160 patterns. Your job is to use this information to build different pattern recognition architectures from the methods that you learned in class over the semester. *Include as many methods as number of members in the group. Note there is now no limit to group size but that the entire group will receive the same grade.*

Submit a brief write-up of each pattern recognition system and include the classification errors as a 2×2 *confusion matrix* matrix (dogs classified as dogs, dogs classified as cats, cats classified as cats and cats classified as dogs). Split your data set into a training and testing data set and provide errors for both. In addition, once you are satisfied with your pattern recognition on the known data, classify the last 38 unknown columns as either cats or dogs. Send your result as a row vector of zeros and ones to me via email.

Here are some examples of how you might proceed:

- Determine the covariance matrix of the cats and the covariance matrix of the dogs and construct optimal bases for each using maximum noise fraction. Project new samples onto the cat basis and dog basis and see which gives a better representation.
- Use vector quantization, e.g., employ Kohonen's self-organizing mapping to see if there is a pattern of distribution on a 2D lattice.
- Use symmetric eigen-cats and eigen-dogs (you will need to ask for the raw data if you do this).

- Wavelets (probably best applied to raw data)
- Radial basis functions (map cats to ones and dogs to zeros).
- Fourier analysis, data frequency content.
- Fisher's linear discriminant analysis.
- Labelled Voronoi cell classification.

Note that our last two class sessions on May 4 and May 6 will be in the computer lab Weber 201. I will answer coding questions and give project assistance. Class attendance is expected.