

Dimension Reduction and Watermarking (Last Homework!)

Homework 6

Due 12/1/2004

1. The goal of this problem is for you to perform feature subset selection. You are given the following datasets:
 - Matlab file `train.mat` containing the following matrices:
 - `x1`: training set (row vectors)
 - `clab1`: training set labels
 - Matlab file `test.mat` contains the following matrices:
 - `x2`: test set (row vectors)
 - `clab2`: test set labels
 - (a) Write a simple K-mean clustering with $K=5$ (or if you want a GMM classification with 5 components) based on the training data. Report the amount of misclassification on the test data `x2`.
 - (b) Perform feature subset selection to determine a reduced number of features that (hopefully) perform better than in the original space. You may employ any feature subset selection technique.
 - (c) Prepare a MATLAB program called `classify.m` that will load `test.mat`, perform the feature selection, and classify each of the examples in the dataset `x2`. Send your `classify.m` to me through email, I will run your program with a separate `test.mat` file containing my own test data. Your grade will be based on the performance of your classifier on my test data. All datasets will obviously be generated from the same distribution.
 - (d) DESCRIBE YOUR APPROACH AND DISCUSS YOUR RESULTS. What search technique(s) did you try? What objective function(s) did you try? What classifier(s) did you employ? Which features were selected? How did you determine how many features to settle for?
2. Evaluate the following watermarking schemes. Consider the approximate density of marks (i.e., how many kilo/megabytes of data are required per mark, and how many marks one could expect to embed in the item in question), the efficiency of the marking and extraction procedures, whether the marks are perceptible (when comparing the marked version to the original), and whether the marks are resistant to detection and/or removal (ignoring the possibility of collusion).
 - In a plain text document containing one of Shakespeare's plays, at the end of each line either a blank space is added (to mark a 0), or no space is added (to mark a 1).
 - In a highresolution, highcolor photograph stored in a nonlossy format, for each pixel, either its blue color component is incremented (to mark a 0), or its red color component is incremented (to mark a 1).
 - In a fulllength motion picture, at every transition from one continuous shot to another, either the last frame of the ending shot is dropped (to mark a 0), or the first frame of the beginning shot is dropped (to mark a 1).

- In the executable file for Powerpoint, every ADD X,Y instruction is either converted to SUB X,Y (to mark a 0), or left alone.
3. Watermarks aren't always used to deter copying. Consider the following scenario: the MediaWare company (which is in both the hardware and media business) plans to freely give away a device that plays music or videos in some standard format. However, MediaWare only wants the device to play media that is sold by MediaWare. Describe how, using watermarking and some other cryptographic primitive we have discussed in this class, MediaWare can design its playback device and mark its media to achieve these goals.
 4. (a) Describe your most favorite and least favorite topics in this class and why.
(b) What do you think the web search engine will be like in ten years?