

CS 4803A/8803A: Pattern Recognition

Problem Set 4

Date: March 13, 2003

Due: March 25, 2003

For the first two problems of this problem set you will compare two different non-parametric classification techniques in the same data set. The training and testing data can be obtained from the class Web page (files PS4_1train.mat and PS4_1test.mat) These are two classes with equal prior probabilities and scalar-valued features. Each classification technique has a free parameter which you will estimate via the technique of *leave-one-out validation* (to be explained).

1. Use the Parzen window technique in DHS 4.3 to estimate the density of each class. Use the Gaussian window (equation (27) in DHS). Unfortunately, the width parameter h is still to be determined for each class.

To estimate h , you will use the amazing technique of leave-one-out validation. The idea is to compute an *evidence curve* $v(h)$ which approximately represents the likelihood of h given the data. Then you can choose the h which maximizes $v(h)$. Each class has its own density so this needs to be done separately for each class. Here's what you do to compute $v(h)$ for a particular value of h :

- (a) Break the training data into two parts A and B , where B contains only a single sample. This is what makes the method "leave-one-out" validation.
- (b) Compute the Parzen window density estimate $p_n(x)$ from A only.
- (c) Compute $\log(p_n(B))$, the log-probability of the sample you held out.
- (d) Iterate back to part (a) and break the training data up in a different way. Do this for every possible choice of B . The average of all of the log-probabilities you get in part (c) is defined to be $v(h)$.

For each class, plot the evidence curve which is a function of h . Pick three values of h from different parts of the curve and plot the estimated density using each. What would your intuition say is the best h ? How well does the evidence curve match your intuition?

Now design a minimum error rate classifier using the h that maximizes $v(h)$ for each class. (Remember, minimum error rate would mean posterior is greatest. With equal priors, simply choose class with greater likelihood - right?) What is its performance on the test data?

2. This time use the k -nearest neighbor method in DHS 4.5.4. Here we do not know the right value of k and will estimate it using an evidence curve $v(k)$. Here's what you do for a particular k :
 - (a) Break the training data into two parts A and B , where B contains only a single sample.

- (b) Determine if B is classified correctly by its k -nearest neighbors from A .
- (c) Iterate back to part (a) and break the training data up in a different way. Do this for every possible choice of B . The number of correct classifications in part (b) is defined to be $v(k)$.

Plot the evidence curve for odd values of k from 1 to 19. Pick three values of k from different parts of the curve and plot the resulting discriminant “function” over the scalar feature space (this function should be 1 when choosing class 1 and -1 otherwise). What would your intuition say is the best k ? Design a minimum error rate classifier using the k that maximizes $v(k)$. What is its performance on the test data?

3. In this problem, you are given high dimensional features derived from samples of $c = 3$ Brodatz textures: marble, pebbles, and paper. The object is to construct quadratic discriminants to distinguish between the three classes. There will be two phases. In the first phase, use Fisher multiple discriminant analysis to project the high dimensional features onto the maximally discriminating $(c-1)$ D space. In the second phase, assume Gaussian statistics and build classifiers between each pair of classes. Then test the classifiers on the input data.

There are features for 45 marble, 63 pebble, and 72 paper samples in in the ASCII files `class1.dat`, `class2.dat`, and `class3.dat` on the class web page dataset section. You can load them into Matlab with e.g. `load class1.dat` (these are not `.mat` files so you load them a bit different. In Matlab you can reference them with e.g. `class1(row,col)`. Each row corresponds to a different sample of the class, and within the row, each of the 30 values is a separate feature (in this case eigenvalues computed on the whole Brodatz texture database).

- (a) Following the method given in the handout from Duda & Hart, use the sample data to construct S_W and S_B . To make a projection matrix \mathbf{W} , you’ll need to find generalized eigenvectors and eigenvalues. Use Matlab’s function `[aa,bb,q,z,v] = qz(a,b)`. Some hints about `qz` are in order: the i^{th} eigenvector appears as column i in `v`; and the corresponding eigenvalue is the ratio `aa(i,i)/bb(i,i)`. Due to tiny numerical errors, Matlab may return some complex numbers. If `imag(z) << real(z)`, just use the real part. Now that you have a projection matrix \mathbf{W} , project the 30D data down to a $c-1 = 2$ D space, and make a scatter plot of that 2D data, with a different symbol for each class.
 - (b) Assume the 2D data is Gaussian and use it to compute individual covariance matrices \mathbf{K}_i for each class. Now use them to make a quadratic classifier for the classes; we want to see the algebraic form of each classifier with numerical values for \mathbf{K}_i and \mathbf{m}_i . Show how well the classifier handles the input data sets. How many samples does it misclassify?
 - (c) Finally, assume all covariances are the same $\mathbf{K} = \mathbf{K}_i$ and make a linear classifier (boy this goes way back in the class). Again, show the algebraic form and how well it works. How much worse is the linear classifier?
4. Consider **both** of the following problem scenarios and consider how pattern recognition may be used to solve it. Compare the two scenarios with respect to the choices you make.

Scenario 1 A cattle rancher wants to have a non-invasive system of recognizing cows by the patterns on their backs. He has a few hundred cows, and they will walk under an archway where a photo is taken of their back from a camera mounted above. You are given only the image data, and you need to tell him the name of this cow. (The rancher wants to keep track of individual data on each cow, their health shots, etc.)

Scenario 2 The post office wants to speed up the delivery of mail by automating the verification of postage stamps on each letter. A camera will snap a photo of each letter as it moves by on a conveyor belt. Given this image, you must find the stamp(s), read their values, and output the amount of postage on the letter. (Assume your output will be put into another system for comparison with the weight of the letter.)

You should consider (but not feel limited to) ideas from matched filtering, eigenvectors, decision theory, and linear/quadratic discriminants. Please limit your response to 4 pages. Make sure your solution includes the following:

- (a) Describe possible features you would collect, and any training processes you might use.
- (b) How will you compare features and make decisions?
- (c) What assumptions are you making?
- (d) How would you test your system? Specify error criteria.
- (e) Name some conditions where you expect your system would fail.
- (f) Suggest some ways these failures might be handled.

Please be as clear as you can, showing mathematically and with diagrams how your solution might work. Feel free to discuss your ideas with others on this problem, but please design your own final system and write your own response.