ECSE 35-6610 Pattern Recognition - Fall 2004

Tuesday & Friday 2:00 - 3:20 Carnegie 102 (?)

Instructor ....Professor George Nagy (nagy@ecse.rpi.edu)
Office hours......................................Tuesday & Thursday 11:00 - 1:00, or by appointment, JEC 6020, 276-6078
Prerequisites ....Probability; programming skill
Grading...5 programming assignments 50%
Term paper (due October 15) 10%
Midterm examination (October 19) 15%
Final (as scheduled by the Registrar) 25%
Text ....Duda, Hart and Stork, *Pattern Classification*, Wiley 2001

Topics

Linear and nonlinear decision procedures
Parametric and nonparametric classifiers
Adaptation and non-supervised learning
Feature extraction and selection
Parameter estimation and performance evaluation (bias and variance)
Clustering, mixture distributions, Expectation Maximization
Linguistic and style context, Hidden Markov Methods
Applications to Optical Character Recognition

All of the programming assignments will be based on a file of optically scanned typeset characters, which we will use to compare the performance of various classification algorithms.

Course objective

On completion of the course, students should be sufficiently familiar with the formal theoretical structure, notation, and vocabulary of pattern recognition to be able to pursue matters of interest in the current technical literature. They will also understand some of the engineering aspects of a prototypical application of pattern recognition, Optical Character Recognition.

Examinations will be open book, but Cheating (copying on examinations or programs) will result in a grade of F for the course. Discussing programming assignments between students is encouraged, but each team must write a separate program, and a separate report. Assignments up to one week late will be given half marks. Assignments more than one week late will not be accepted. Special considerations will, however, be given to special situations, such as the DQE: they must be negotiated in advance. Grades may be appealed initially to Professor Nagy: if the difficulty cannot be resolved, procedures in the Rensselaer Handbook must be followed.

35-661 Assignments and Due Dates

All assignments are due at the beginning of the class!

1 Moment feature extraction from printed characters...09-17
2 Bayes classifier for binary and Gaussian distributions 10-1
3 Term paper 10-15
MIDTERM **10-19**
4 Nearest-neighbors classifier 10-29
5 Decision tree classifier 11-12
6 Clustering and unsupervised learning...12-3
FINAL **scheduled by the Registrar**

ASSIGNMENT #1 - DUE Monday, September 17, 2004, 2pm.

The objective of this assignment is to extract some continuous-valued features from the scanned characters in our data set. In subsequent assignments we will use them to illustrate algorithms based on normal distributions.

Each character is viewed as a 2-D binary array. Let **i be the row index** increasing from *bottom to top of each character*, and **j be the column index** increasing from *left to right of the character*. The origin is fixed so that the first row (i=1) and the first column (j=1) both contain at least one non-zero entry. **It is important that everyone use the same coordinate system, otherwise we cannot compare results!** You are requested to test your program on some standard data (in ascii format) on **/home/10/nagyg/public/pattern-recognition**. For this assignment use only **Dataset A.** Note that in the file the bottom row of the character is listed first, so the characters appear upside down. Print out all the characters as an image with the correct aspect ratio to make sure that you are reading the data right, and to identify the character labels. Add coordinate axes so that one can identify pixel (i,j)..

We will use ***k* as the class index**, with **k=1 (not 0!) for "a" and k=10 for "z".**
The characters in each class are also labeled 1 to 10: a1, a2, ...

Definition: The (p,q)th Central Moment of a pattern X is:

$$Mpq = \Sigma\Sigma\ [((i-imean)^p)\ ((j-jmean)^q)\ X(i,j)], \qquad i=1 \text{ to } I,\ j=1 \text{ to } J.$$

where X(i,j) is the pixel (1 or 0) in the (i,j)th entry of the pattern matrix, I is the number of rows, J is the number of columns, and (imean, jmean) is the centroid of the pattern. The values of imean and jmean should not be rounded or truncated.

1. For the 10 samples of each class of the training set, compute 8 central moments of the class-conditional sample distributions for each class. The moments vary greatly in magnitude, and some are negative, **so normalize the moments by dividing the moments of each pattern by the overall (all 100 samples) root-mean-square value for that moment (*overall rms*)**. For each class, compute the class mean and class variance of the *normalized* moments. (To compute the variance, divide by N rather than by N-1.)
Print a table **in fixed-point format (rows, columns, and decimal points aligned)** showing the feature values and class-conditional statistics as shown on the next page.

2. Compute the *class-conditional sample covariance matrix* ^kC for each class k in the training set. Divide by N throughout, rather than N-1. Print these matrices for Dataset A and B. Please label each matrix **unambiguously**.as **Cov-Dataset-A-a;... Cov-Dataset-A-z.** You may use any software package for computing the covariance matrices, or write your own. Please make sure that matrices don't span page breaks.

3. Invert each matrix and print ^kC, $^kC^{-1}$ (k = 1, 2, ...,10) as 8x8 arrays.
Check that $^kC\ ^kC^{-1}$ = I. (For debugging, you may want to work with the 2x2 covariance matrices of the first two features.) Please print all matrices in **fixed-point format (rows, columns, and decimal points aligned)** for legibility. Make sure that you normalize your values by the appropriate power of 10 to show 2 or 3 significant digits. Label as **Inverse-Cov-Dataset-A**, etc. Show the multiplier (*true value = multiplier x value printed*).

4. Compute and print the *average covariance matrix C* by averaging the ten class-conditional matrices term by term. Compute its *inverse* C^{-1} and label it **Inverse-Cov-Dataset-A**, etc.

Formatting the data in a sensible manner is an important part of this homework.
The features will have to be computed for other datasets as well, so be sure to document and save your programs. **Please hand in a copy of your source code in a separate package.**

Normalized Central Moments (Dataset-A)

	M_{00}	M_{02}	M_{11}	M_{20}	M_{03}	M_{12}	M_{21}	M_{30}
a1	0.908	0.307	...					
a2	0.946	0.318	...					
.								
.								
a10								
mean(a)	1.002	0.357						
var (a)	0.003	0.001						
c1	0.687	0.302						
c2	0.709	0.291						
.								
z10								
mean(z)								
var (z)								
overall rms	131.09	3797.28						

Store the calculated values, round-off only for printing!

The value of the (1,1) entry for Cov-Dataset-A-a is 0.002701…
The value of the (1,1) entry of its inverse is 54530.

Dataset-A:

a a a a a a a a a a

c c c c c c c c c c

e e e e e e e e e e

m m m m m m m m m m

n n n n n n n n n n

o o o o o o o o o o

r r r r r r r r r r

s s s s s s s s s s

x x x x x x x x x x

z z z z z z z z z z

Dataset-B:

a a a a a a a a a a

c c c c c c c c c c

e e e e e e e e e e

m m m m m m m m m m

n n n n n n n n n n

o o o o o o o o o o

r r r r r r r r r r

s s s s s s s s s s

x x x x x x x x x x

z z z z z z z z z z

Assignment #3 - Term Paper - Due Friday October 15, 2004

The objectives of this assignment are to familiarize you with pattern recognition periodicals; to expose you to a topic possibly not covered in class; to give you practice in preparing a paper according to standards expected in submission for publication. The periodicals published by the professional societies are usually more thoroughly refereed than conference papers or proclamations posted on the Web. Of course, the full text of many recently published articles are available on the Web.

Find three articles, by distinct authors, on a **single topic** in statistical pattern recognition, in:

> ***IEEE Transactions on Pattern Analysis and Machine Intelligence*** (*best*)
>
> For a list of other technical journals, please see:
> http://www.ph.tn.tudelft.nl/PRInfo/journals.home.html

Prepare a 4-6 page (double-spaced) critique of the subject that you have chosen. Integrate the findings of the three papers and comment on how they support or negate each other's observations. Evaluate whether their main conclusions are credible.

In addition to your choice of articles and the perceptiveness and incisiveness of your observations, your paper will be graded on style, including organization, syntax, and professional presentation. Write concisely and precisely. Be sure to provide the *complete* bibliographic citations of the papers you review (*use the same format for citations as one of the journals from which you take your material*), and **please append copies of the papers** (they will be returned).

The best papers will be circulated to the class. You are urged to ask others, including classmates, to read your drafts and make suggestions. Your message should be understandable without reference to the original articles.

If you wish to use material from other sources, please check with me by October 2. Avoid conference proceedings, which often contain hastily assembled and incomplete reports.
I shall also be glad to comment on a draft if you get it to me at least two weeks before the due date, but this is optional. **Start now**.

Style guides:

C. Waddell, BASIC PROSE STYLE AND MECHANICS, copies available from the RPI Writing Center, 1987.

R. A. Day, HOW TO WRITE AND PUBLISH A SCIENTIFIC PAPER, ISI Press, Philadelphia, PA, 1979.

R. S. Smith, WRITTEN COMMUNICATION FOR DATA PROCESSING, Van Nostrand, 1976.

K.K. Landes, A SCRUTINY OF THE ABSTRACT, Bull. of the Am. Ass. of Petroleum Geologists 35, 7, pp. 1660, July 1951.

Have Professor Nagy sign your bibliography by Friday, October 1.

SAMPLE TEST

____
name

PATERN RECOGNITION MIDTERM

October 19, 1858 Open Book

Please copy all your answers in the blanks on this sheet, but show and hand-in all your work

1. We classify objects into two classes A and B using three binary features x, y, and z and a minimum-error classifier. The features may be assumed to be class-conditionally statistically independent. The statistics derived from the training sample are:

P(A) = 0.8

Conditional probability of each feature having value "1":

	x	y	z
class A	.3	.5	.9
class B	.6	.5	.3

a. (1 pt) How will it classify the pattern "1 1 0" ? ______________

b. (2 pts) What is the probability of error **given** that the measurement vector is "1 1 0" ? ______________

c. (2 pts) What is error probability of this classifier? ______________

d. (1 pt) Give an optimal linear discriminant of the form
ax + by + cz + d > 0 iff (x,y,z|A) > P(x,y,z|B).

a = ______ b = ______ c = ______ d = ________

2. Perhaps for the last time in your life, you are asked to estimate the bias h of a coin (h is the probability of *heads*). Suppose that after ten tosses $x_1, \ldots, x_{10}$, the probability density of h given the ten samples is:

$$p(h \mid x_1, x_2, \ldots, x_{10}) = 12\, h^3 (1 - h^2)$$

a. (4 pts) If the next toss yields *heads*, what will be the Bayes estimate of h? h_B = ______

b. (2 pts) If all the odd-numbered tosses were heads,
what will be the Maximum Likelihood estimate? h_{ML} = _____

3. We observe pairs of printed digits. Only the digits "0" and "1" can occur.
Two binary-valued measurements x and y are taken on each digit.
The measurements on each digit depend only on whether that digit is "0" or "1" (*not* on its position). The measurements are class-conditionally independent of each other with the following probabilities of having value 1.

	x	y
digit "0"	0.1	0.2
digit "1"	0.8	0.9

The *a priori* probabilities of the pairs of digits are:

P["00"] = 0.2
P["01"] = 0.3
P["10"] = 0.4
P["11"] = 0.1

(3 pts) Suppose that we observe a pair of digits, and all four measurements have value **one**.
What is the probability that the **second** digit is "**0**"?