CAP 6638 **Pattern Recognition** **Spring 2015**

MW 05:00 PM – 06:15 PM in NES 013

Project #2

Please program the six different trainable classifiers listed below and report the classification results. Train each classifier on Data Set A, and use it to classify Date Set A and the three test sets B, C, and D. (Assign ties to "smallest first"): For normalization, use the rms values *from Data Set A* for each test set.

1. *Minimum distance moment classifier*: each sample is assigned to the category whose class mean (centroid) is nearest according to the Euclidian distance in the eight-dimensional measurement space. This is the same as Bayesian classification with *identity* covariance matrices.

2. *Bayes moment classifier with identical covariances*: use the average covariance matrix. This results in a linear classifier.

3. *Bayes moment classifier with individual class covariances*: quadratic decision function.

4. Same as (3), but using only *the first four moments.*

5. *Minimum distance classifier in binary pixel space.* For the class centroids take the Bayes estimate of the pixel probabilities, with uniform *a priori* density (i.e., (m+1)/(n+2) where m is the number of 1s at a pixel for the class and n is the total number of samples per class (10 in our case)). Use Euclidian distance.

6. *Bayes classifer in binary pixel space*, assuming class-conditionally independent measurements. Use the Bayes estimate of the pixel probabilities, with uniform *a priori* density as in (5).

At the beginning of your report, summarize your error rates in a well-formatted table of the error rates, with one row for each method. Follow this with a one or two page summary showing the equations as you used them, and comment on your results. Next, print the score tables and confusion tables, **sorted first by Method**, next by Data Set. Print each confusion table (for Method 1 only) right after the corresponding score table. Examples of the other tables are given on the page. The values are hypothetical.

SUMMARY TABLE

Error counts on six Bayesian classifiers

Test set: Method	**A**	**B**	**C**	**D**
1	1	2	1	0
2	0	3	2	0
3	0	4	4	2
4				
5				
6				

For each classifier, but only for Data Set B, tabulate the 10 class scores for each character, using one line per character in the format below (hypothetical results). **Use fixed-point format so that the scores for each class form columns. Use scale factors if necessary to avoid decimal points. Avoid page breaks within an array**. Title each table carefully. Mark incorrectly classified samples on the right as shown: Print the sample number (**starting with 1, not 0**) in the first column, so that we can conveniently compare results.

The values given here and the summary table above are random numbers, please do not compare it with the value you get in your experiments.

Method 1 – Eight Moments, Identity Covariance Matrix **(trained on A, tested on B)**

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**0**	**ERRORS**
1-1	234	213	211	145	120	133	145	229	150	220	
1-2	236	222	234	146	130	135	150	238	220	211	**u**

1-10 .....

3-1 .....
3-2 .....

0-10 .....

Method 1 – Eight Moments, Identity Covariance Matrix **(trained on A, tested on B)**

1-1 .....
1-2 .....

For the first classifier only, for each dataset, print out a *confusion table* modeled on the following *hypothetical* example:

Confusion Table
Method 1 – Eight Moments, Identity Covariance Matrix (trained on A, tested on A)

Classified as:	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**0**	Error
True class											Type I
1	9					1					1
2		8							2		2
3			8						1	1	2
4				9						1	1
5					10						0
6						10					0
7							10				0
8								10			0
9									7	3	3
0										10	0
Error Type II	0	0	0	0	0	1	0	0	3	5	**9**

Each entry represents the number of samples (an integer) of a particular class classified in particular way, with the actual class corresponding to the row and the decision corresponding to the column. In the bottom right corner, print the total number of errors. Type I Error for "1" is the number of "1"s that are misclassified. Type II Error for "1" is the number of times a digit is misclassified as "1".

There will be altogether **6** score tables and **4** confusion tables. Formatting may be a lot of work, but the results of experiments must be readily understood.

Submission materials: Source code and report in Canvas. Create a zip file including all materials.