ICS 663: Pattern Recognition (Fall 2015)

Homework Assignment #1
(Due: **5:00 PM, Monday September 21**)

- When answering questions, show how you arrived at the answer, without missing crucial steps in reasoning.
- To help the grader (and also yourself), please be neat, document your code properly, and answer the questions in the order they are stated.
- Typed work is preferred. But if you submit your work written by hand, make your handwriting as clear and recognizable as possible. You are solely responsible for any consequences resulting from poor handwriting.
- If you submit your homework either electronically (using the Drop Box folder in the course website at Laulima) or in person. If you submit it in person, staple all pages together and be sure to write your name on the front page.
- Programming assignments:
  - Make a single pdf file of report describing
    - ✓ implementation detail,
    - ✓ how to run your code including necessary software or system set-up,
    - ✓ results and/or answers to the questions.
  - **Upload your code (including executable file(s)) and the report** in your Drop Box folder at Laulima. In the code, make sure you type in your name and any necessary comments.
  - You may use any built-in functions to calculate mean, covariance and matrix determinant.

1. In many pattern classification problems one has the option either to assign the pattern to one of c classes, or to *reject* it as being unrecognizable. If the cost for rejects is not too high, rejection may be a desirable action. Let

$$\lambda(\alpha_i \mid \omega_j) = \begin{cases} 0 & i = j \quad i, j = 1, \ldots, c \\ \lambda_r & i = c+1 \quad \text{(rejection)} \\ \lambda_s & \text{otherwise} \end{cases}$$

where λ_r is the loss incurred for choosing the $(c + 1)^{\text{th}}$ action, rejection, and λ_s is the loss incurred for making any substitution error.

(a) (**10 points**) Show that the minimum risk is obtained by the following decision rule.

$$\text{Decide } \begin{cases} \omega_i & \text{if } P(\omega_i \mid \mathbf{x}) \geq P(\omega_j \mid \mathbf{x}) \text{ for all } j, \text{and if } P(\omega_i \mid \mathbf{x}) \geq 1 - \dfrac{\lambda_r}{\lambda_s} \\ \textit{reject} & \text{otherwise} \end{cases}$$

(b) (**5 points**) Describe what happens if $\lambda_r = 0$.

(c) (**5 points**) Describe what happens if $\lambda_r > \lambda_s$.

(d) (**10 points**) Show that the following discriminant functions are optimal for this problem.

$$g_i(\mathbf{x}) = \begin{cases} p(\mathbf{x} \mid \omega_i) P(\omega_i) & i = 1, \ldots c \\ \dfrac{\lambda_s - \lambda_r}{\lambda_s} \displaystyle\sum_{j=1}^{c} p(\mathbf{x} \mid \omega_j) P(\omega_j) & i = c+1 \end{cases}$$

(e) (**7 points**) Plot these discriminant functions and the decision regions for the two-category one-dimensional case having the following. (You can use any software for plotting.)

- $p(x \mid \omega_1) \sim N(1,1)$
- $p(x \mid \omega_2) \sim N(-1,1)$
- $P(\omega_1) = P(\omega_2) = 0.5$
- $\dfrac{\lambda_r}{\lambda_s} = \dfrac{1}{4}$

2. Let the components of the vector $\mathbf{x} = (x_1,\ldots, x_d)^t$ be binary-valued (0 or 1) and let $P(\omega_j)$ be the prior probability for the category ω_j and $j = 1,\ldots, c$. Now define

$$p_{ij} = P(x_i = 1|\omega_j) \text{ for } i = 1,\ldots, d \text{ and } j = 1,\ldots, c,$$

with the component of x_i being statistically independent for all $\mathbf{x}$ in ω_j.

   (a) (**5 points**) Interpret in words the meaning of p_{ij}.

   (b) (**8 points**) Show that the minimum probability of error is achieved by the following decision rule: Decide ω_k if $g_k(\mathbf{x}) \geq g_j(\mathbf{x})$ for all j and k, where

$$g_j(\mathbf{x}) = \sum_{i=1}^{d} x_i \ln \frac{p_{ij}}{1-p_{ij}} + \sum_{i=1}^{d} \ln(1-p_{ij}) + \ln P(\omega_j)$$

3. (**10 points**) Assume that the samples are drawn independently from a univariate Gaussian distribution. Derive the following equations for the maximum-likelihood estimation of the mean, μ_{ML}, and the variance, σ_{ML}^2.

$$\mu_{ML} = \frac{1}{n}\sum_{k=1}^{n} x_k$$

$$\sigma_{ML}^2 = \frac{1}{n}\sum_{k=1}^{n} (x_k - \mu_{ML})^2$$

Programming Exercise: Bayesian Classifier with Normal Distributions

Consider a three-category classification problem with two-dimensional feature vector $\mathbf{x} = (x_1, x_2)$. Assuming that the underlying distribution of each category is normal, 300 samples – 100 per category – were drawn for training. The file '*hw1_traindata.txt*' contains 300 lines corresponding to those training samples. Lines 1 ~ 100 are samples from category 1, 101 ~ 200 are from category 2, and 201 ~ 300 belongs to category 3. The mean vectors and covariance matrices below were used for the data generation.

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{\mu}_2 = \begin{bmatrix} 3 \\ 3 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 1 & 1.6 \\ 1.6 & 4 \end{bmatrix}, \quad \boldsymbol{\mu}_3 = \begin{bmatrix} 0 \\ 3.5 \end{bmatrix}, \Sigma_3 = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

Since the covariance matrices are different for each category, it would be appropriate to use the discriminant function from Case 3 in the lecture notes. However, in this assignment you will estimate the performance using the discriminant functions from all three cases. In training, use the mean vectors and covariance matrices estimated from the training data.

Training: Suppose a uniform prior probability for the categories. Classify the training samples using the discriminant functions obtained by assuming:

- Case 1: $\Sigma_{\boldsymbol{i}} = \sigma^2\mathbf{I}$ for *i* = 1, 2, 3. Assume unit variance, i.e. $\sigma^2 = 1$.
- Case 2: $\Sigma_{\boldsymbol{i}} = \Sigma$ for *i* = 1, 2, 3. In this case, use the average covariance matrix as the common covariance matrix.
- Case 3: $\Sigma_{\boldsymbol{i}}$ is arbitrary.

(1) Make a scatterplot of the training samples. Indicate the estimated mean of each category. On top of the scatterplot, show the decision boundaries obtained for each of the cases above.

(2) Report the training error rate for each of the cases above.

Testing: The file '*hw1_testdata.txt*' contains 300,000 lines corresponding to those test samples. Again, the first 1/3 of the samples is from category 1, the middle 1/3 from category 2, and the last 1/3 belongs to category 3. Classify these samples using the three discriminant functions learned above, and report the error rate for each of the cases above.

<u>Grading for programming exercises</u>:

- 5 points – Correct scatterplot of the training samples with mean of each category indicated
- 15 points – Correct decision boundaries (5 points for each case)
- 7.5 points – Correct training error rates (2.5 points for each case)
- 7.5 points – Correct test error rates (2.5 points for each case)
- 5 points – Report

* If your code does not work you will receive credits for your report only.