Problem Set 2

MAS 622J/1.126J: Pattern Recognition and Analysis

Due Monday, 2 October 2006

[Note: All instructions to plot data or write a program should be carried out using either Python accompanied by the `matplotlib` package or Matlab. Feel free to use either or both, but in order to maintain a reasonable level of consistency and simplicity we ask that you do not use other software tools. Please provide a printed copy of any code used for this problem set.]

Problem 1:

In a particular binary hypothesis testing application, the conditional density for a scalar feature y given class w_1 is

$$p_{y|w_1}(y|w_1) = k_1(exp(-y^2/10))$$

Given class w_2 the conditional density is

$$p_{y|w_2}(y|w_2) = k_2(exp(-(y-2)^2/2))$$

a. Find k_1 and k_2, and plot the two densities on a single graph using Matlab/Python.

b. Assume that the prior probabilities of the two classes are equal, and that the cost for choosing correctly is zero. If the costs for choosing incorrectly are $C_{12} = 1$ and $C_{21} = \sqrt{5}$, what is the expression for the Bayes risk?

c. Find the decision regions which minimize the Bayes risk, and indicate them on the plot you made in part (a)

d. For the decision regions in part (c), what is the numerical value of the Bayes risk?

Problem 2:

[Note: Use Matlab or Python for the computations, but make sure to explicitly construct every transformation required, that is either type it or write it. Do not use Matlab or Python if you are asked to explain/show something.]

Consider the three-dimensional normal distribution $p(\mathbf{x}|w)$ with mean μ and covariance matrix $\mathbf{\Sigma}$ where

$$\mu = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \mathbf{\Sigma} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 6 & 3 \\ 0 & 3 & 6 \end{pmatrix}.$$

Compute the matrices representing the eigenvectors and eigenvalues Φ and Λ to answer the following:

a. Find the probability density at the point $\mathbf{x_0} = (.75\ 0\ 1)^T$

b. Construct an orthonormal transformation $\mathbf{y} = \Phi^T\mathbf{x}$. Show that for orthonormal transformations, Euclidean distances are preserved (i.e., $\|y\|^2 = \|x\|^2$).

c. After applying the orthonormal transformation add another transformation $\Lambda^{-1/2}$ and convert the distribution to one centered on the origin with covariance matrix equal to the identity matrix. Show that $\mathbf{A_w} = \mathbf{\Phi\Lambda^{-1/2}}$ is a linear transformation (i.e., $\mathbf{A_w}(a\mathbf{x} + b\mathbf{y}) = a\mathbf{A_w x} + b\mathbf{A_w y}$)

d. Apply the same overall transformation to $\mathbf{x_0}$ to yield a transformed point $\mathbf{x_w}$

e. Calculate the Mahalanobis distance from $\mathbf{x_0}$ to the mean μ and from $\mathbf{x_w}$ to $\mathbf{0}$. Are they different or are they the same? Why?

f. Does the probability density remain unchanged under a general linear transformation? In other words, is $\mathbf{p}(\mathbf{x_0}|\mu, \mathbf{\Sigma}) = \mathbf{p}(\mathbf{Z^T x_0}|\mathbf{Z^T}\mu, \mathbf{Z^T\Sigma Z})$ for some linear transform $\mathbf{Z}$? Explain.

Problem 3:

Let $\mathbf{x}$ be an observation vector. You would like to determine whether $\mathbf{x}$ belongs to w_1 or w_2 based on the following decision rule, namely *decision rule 1.*

> Decide w_1 if $-\ln \mathbf{p}(\mathbf{x}|w_1) + \ln \mathbf{p}(\mathbf{x}|w_2) < \ln\{\mathbf{P}(w_1)/\mathbf{P}(w_2)\}$; otherwise decide w_2.

You know that this rule does not lead to perfect classification therefore you must calculate the probability of error. Let χ_1 and χ_2 be the region in the domain of $\mathbf{x}$ such that

> $\mathbf{p}(\mathbf{x}|w_1)\mathbf{P}(w_1) > \mathbf{p}(\mathbf{x}|w_2)\mathbf{P}(w_2)$ and $\mathbf{p}(\mathbf{x}|w_1)\mathbf{P}(w_1) < \mathbf{p}(\mathbf{x}|w_2)\mathbf{P}(w_2)$, respectively.

Then if $\mathbf{x} \in \chi_i$, for $i = 1, 2$ assign the sample to class w_i. Use excruciating detail to answer the following:

a. Show that the Pr[error] for this rule is given by:

$$\Pr[\text{error}] = \mathbf{P}(w_1)\epsilon_1 + \mathbf{P}(w_2)\epsilon_2$$

$$\text{where } \epsilon_1 = \int_{\chi_2} \mathbf{p}(\mathbf{x}|w_1)\mathbf{dx} \text{ and } \epsilon_2 = \int_{\chi_1} \mathbf{p}(\mathbf{x}|w_2)\mathbf{dx}$$

b. Describe what the previous equation says about the total error. (hint: identify what ϵ_1 and ϵ_2 mean)

c. Suppose that for a given decision, you must pay a cost depending on the true class of the sample based on *decision rule 1.* Assume that a wrong decision is more expensive than a correct one, where $\lambda_{ij} = \lambda($deciding $w_i|w_j)$ is the loss incurrred for desiding w_i when the state of nature is w_j. Write an expression for the expected cost, namely risk, R, such that

$$E[\text{cost}] = E[\text{fixed costs}] + E[\text{variable costs}]$$

(hint: recall that the decision rule induces a partitioning of the measurement space into a number of disjoint regions)

d. Suppose that for a given value of $\mathbf{x}$, the integrand in the risk function is positive. How can you decrease the risk? (hint: think about where you would assign $\mathbf{x}$ to and why you would make that decision)

e. Show that for a symmetrical cost function $\lambda_{12} - \lambda_{22} = \lambda_{21} - \lambda_{11}$, $E[\text{cost}] = \Pr[\text{error}]$.

Problem 4:

Use signal detection theory as well as the notation and basic Gaussian assumptions described in the text to address the following

a. Prove that $\mathbf{P}(\mathbf{x} > \mathbf{x}^*|\mathbf{x} \in w_2)$ and $\mathbf{P}(\mathbf{x} > \mathbf{x}^*|\mathbf{x} \in w_1)$, taken together, uniquely determine the discriminability $\mathbf{d}'$

b. Use error functions $erf(*)$ to express $\mathbf{d}'$ in terms of the hit and false alarm rates. Estimate $\mathbf{d}'$ if $\mathbf{P}(\mathbf{x} > \mathbf{x}^*|\mathbf{x} \in w_1) = .6$ and $\mathbf{P}(\mathbf{x} > \mathbf{x}^*|\mathbf{x} \in w_2) = .5$. Repeat for $\mathbf{d}'$ if $\mathbf{P}(\mathbf{x} > \mathbf{x}^*|\mathbf{x} \in w_1) = .9$ and $\mathbf{P}(\mathbf{x} > \mathbf{x}^*|\mathbf{x} \in w_2) = .2$.

c. Given that the Gaussian assumption is valid, calculate the Bayes error for both the cases in (b).

d. Using a trivial one-line computation or a graph determine which case has the higher $\mathbf{d}'$, and explain your logic:

   Case A: $\mathbf{P}(\mathbf{x} > \mathbf{x}^*|\mathbf{x} \in w_1) = .7$ and $\mathbf{P}(\mathbf{x} > \mathbf{x}^*|\mathbf{x} \in w_2) = .4$.

   Case B: $\mathbf{P}(\mathbf{x} > \mathbf{x}^*|\mathbf{x} \in w_1) = .2$ and $\mathbf{P}(\mathbf{x} > \mathbf{x}^*|\mathbf{x} \in w_2) = .8$.

Problem 5:

a. Show that the maximum likelihood (ML) estimation of the mean for a Gaussian is unbiased but the ML estimate of variance is biased (i.e., slightly wrong). Show how to correct this variance estimate so that it is unbiased.

b. For this part you'll write a program with Matlab/Python to explore the biased and unbiased ML estimations of variance for a Gaussian distribution. Find the data for this problem on the class webpage as ps2.dat. This file contains n=5000 samples from a 1-dimensional Gaussian distribution.

   (a) Write a program to calculate the ML estimate of the mean, and report the output.

   (b) Write a program to calculate both the biased and unbiased ML estimate of the variance of this distribution. For n=1 to 5000, plot the biased and unbiased estimates of the variance of this Gaussian. This is as if you are being given these samples sequentially, and each time you get a new sample you are asked to re-evaluate your estimate of the variance. Give some interpretation of your plot.

Problem 6:

DHS 3.7