MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Media Arts and Sciences Section

Course 4.902: Pattern Recognition and Analysis — Assigned: 2/16/93

Problem Set No. 3 — Due: 2/23/93

READING ASSIGNMENT: Therrien, Sections 4.5-5.3

This problem set consists of seven problems. Remember that you are *not* allowed to look at old problem sets, but collaboration is allowed and encouraged.

Problem 1: (15 points)

In a particular binary hypothesis testing problem, the decision as to which class a scalar observation $\hat{y}$ belongs to is made by comparing the likelihood ratio $l(\hat{y})$ with a threshold λ, as follows.

$$l(\hat{y}) = \frac{p_{y|\omega_1}(\hat{y}|\omega_1)}{p_{y|\omega_2}(\hat{y}|\omega_2)} \quad \begin{matrix} \omega_1 \\ > \\ < \\ \omega_2 \end{matrix} \quad \lambda$$

a) Show that $E[\ln l(y)|\omega_1] \geq 0$, where

$$E[\ln l(y)|\omega_1] = \int_{-\infty}^{\infty} \ln l(y) p_{y|\omega_1}(y|\omega_1) dy.$$

Hint: Use the fact that $\ln x \leq x - 1$ for all $x > 0$.

b) When is $E[\ln l(y)|\omega_1] \geq 0$ satisfied with equality?

c) Consider the space of all probability density functions for y. Does $E[\ln l(y)|\omega_1]$ satisfy the necessary conditions to be a distance function for the two densities $p_{y|\omega_1}(y|\omega_1)$ and $p_{y|\omega_2}(y|\omega_2)$? If not, which condition(s) does it fail to satisfy?

Note: For two densities $p_x(\hat{x})$ and $q_x(\hat{x})$, the quantity $\int_{-\infty}^{\infty} p_x(x) \ln \frac{p_x(x)}{q_x(x)} dx$ is variously known as the *information divergence*, *relative entropy*, and *Kullback-Liebler distance* between p and q. See *Elements of Information Theory* by T.M. Cover and J.A. Thomas (Wiley, 1991).

Problem 2: (10 points)
A covariance matrix is

$$K = E[(\mathbf{x}-\mathbf{m})(\mathbf{x}-\mathbf{m})^{\mathrm{T}}] = \begin{bmatrix} 5/4 & -\sqrt{3}/4 \\ -\sqrt{3}/4 & 7/4 \end{bmatrix}$$

a) What are its eigenvalues and eigenvectors?

b) Sketch the ellipse representing this covariance matrix in the $\mathbf{x}$ plane.

Problem 3: (10 points)
A random variable x has the probability density

$$p_x(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-\frac{(x-\mu)^2}{2\sigma^2}]$$

a) Find the density of y if $y = ax + b$.

b) Find the density of y if $y = x^2$.

Problem 4: (20 points)
The means and covariances for two classes of Gaussian observation vectors are

$$\mathbf{m}_1 = [1 \;\; 0]^{\mathrm{T}} \qquad \mathbf{m}_2 = [0 \;\; 1]^{\mathrm{T}} \qquad \mathbf{K}_1 = \mathbf{K}_2 = \begin{bmatrix} 1 & -1/2 \\ -1/2 & 1 \end{bmatrix}$$

The prior probabilities of the classes are $\Pr[\omega_1] = 2/3$ and $\Pr[\omega_2] = 1/3$.

a) What is the decision rule that minimizes the probability of error?

b) Let h be the random variable that represents the value of the natural logarithm of the likelihood ratio. Determine the density $p_{h|\omega_1}(\hat{h}|\omega_1)$ and express the error $\epsilon_1 = \Pr[\mathrm{error}|\omega_1]$ as an integral of the density $p_{h|\omega_1}(\hat{h}|\omega_1)$.

Problem 5: (15 points)
A vector $\mathbf{x}$ has the following correlation matrix.

$$\mathbf{R} = \begin{bmatrix} 3 & -2 & 0 \\ -2 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix}.$$

a) Find the eigenvalues and eigenvectors of $\mathbf{R}$.

b) We wish to approximate $\mathbf{x}$ by $\tilde{\mathbf{x}} = c_1\mathbf{e}_1$, where $\mathbf{e}_1$ is the eigenvector corresponding to the largest eigenvalue. Find the expression for c_1 that minimizes the mean-square error between $\mathbf{x}$ and $\tilde{\mathbf{x}}$. What is the value of the resulting minimum mean-square error?

c) Suppose now that we wish to approximate $\mathbf{x}$ by $\tilde{\mathbf{x}} = c_1\mathbf{e}_1 + c_2\mathbf{e}_2$, where $\mathbf{e}_1$ and $\mathbf{e}_2$ are the eigenvectors corresponding to the largest two eigenvalues. Find expressions for the best choices of c_1 and c_2, and determine the value of the corresponding mean-square error.

Problem 6: (15 points)

Consider a vector of integer data $[x_1, x_2, \ldots, x_N]^{\mathrm{T}}$, where N is very large, and where each x_i assumes one of the values in $\{0, 1, \ldots, K-1\}$. For a fixed integer displacement d, we can count the number of times that the joint event $[x_n = i \text{ and } x_{n-d} = j]$ occurs, where $n, n-d \in \{1, \ldots, N\}$ and $i, j \in \{0, 1, \ldots, K-1\}$. If we do this for a fixed d and every possible pair (i, j), we obtain a $K \times K$ matrix $\mathbf{C}_d$ of normalized counts $c_d(i, j)$, which approximate $\Pr[x_n = i \mathrm{and} x_{n-d} = j]$. Such a matrix is called a *co-occurrence matrix.*

Co-occurrence matrices are often used to provide features for classification in image processing. For typical values of K, e.g. $K = 256$ in digitized images, far too many features would result if all of the elements of the matrix were taken as features. It is common therefore to obtain a small set of features by taking scalar functions of the matrix. One such function which is popular in texture classification is:

$$f(d) = N \sum_{i=0}^{K-1} \sum_{j=0}^{K-1} ij c_d(i, j).$$

How does the resulting feature $f(d)$ relate to the autocorrelation of the original vector $[x_1, x_2, \ldots, x_N]^{\mathrm{T}}$? You can assume that $N >> K$ and that the x_n take on values in $\{0, 1, \ldots, K-1\}$ with equal likelihood.

It has been stated in the literature that "co-occurrences are more powerful than correlation for discriminating patterns in images." Comment on this statement in light of the relationship you have just found.

Problem 7: (15 points)

An n-dimensional Gaussian random vector $\mathbf{x}$ has mean $\mathbf{m}$ and covariance matrix $\mathbf{K}$.

a) Show that

$$E[(\mathbf{x} - \mathbf{m})^{\mathrm{T}} \mathbf{K} (\mathbf{x} - \mathbf{m})] = n.$$

b) The entropy H of $\mathbf{x}$ is defined as

$$H = E[\ln p_{\mathbf{x}}(\hat{\mathbf{x}})].$$

Using the property given in part (a), show that

$$H = \frac{1}{2} \ln |\mathbf{K}| + \frac{n}{2}(1 + \ln 2\pi).$$