Problem Set 2

MAS 622J/1.126J: Pattern Recognition and Analysis

Due: 5:00 p.m. on September 30

[Note: All instructions to plot data or write a program should be carried out using Matlab. In order to maintain a reasonable level of consistency and simplicity we ask that you do not use other software tools.]

If you collaborated with other members of the class, please write their names at the end of the assignment. Moreover, you will need to write and sign the following statement: "In preparing my solutions, I did not look at any old homeworks, copy anybody's answers or let them copy mine."

Problem 1: [10 Points]

In many pattern classification problems we have the option to either assign a pattern to one of c classes, or reject it as unrecognizable - if the cost to reject is not too high. Let the cost of classification be defined as:

$$\lambda(\omega_i|\omega_j) = \begin{cases} 0 & \text{if } \omega_i = \omega_j, \text{ (i.e. Correct Classification)} \\ \lambda_r & \text{if } \omega_i = \omega_0, \text{ (i.e. Rejection)} \\ \lambda_s & \text{Otherwise, (i.e. Substitution Error)} \end{cases}$$

Show that for the minimum risk classification, the decision rule should associate a test vector $\boldsymbol{x}$ with class ω_i, if $\mathrm{P}(\omega_i|\boldsymbol{x}) \geq \mathrm{P}(\omega_j|\boldsymbol{x})$ for all j and and $\mathrm{P}(\omega_i \mid \boldsymbol{x}) \geq 1 - \frac{\lambda_r}{\lambda_s}$, and reject otherwise.

Problem 2: [16 points]

In a particular binary hypothesis testing application, the conditional density for a scalar feature $\mathbf{x}$ given class w_1 is

$$p(\mathbf{x}|w_1) = k_1 \exp(-\mathbf{x}^2/20)$$

Given class w_2 the conditional density is

$$p(\mathbf{x}|w_2) = k_2 \exp(-(\mathbf{x}-6)^2/12)$$

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

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 $C_{12} = \sqrt{3}$ and $C_{21} = \sqrt{5}$ (where C_{ij} corresponds to predicting class i when it belongs to class j), what is the expression for the conditional 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 3: [16 points]

Let's consider a simple communication system. The transmitter sends out messages $m = 0$ or $m = 1$, occurring with a priori probabilities $\frac{3}{4}$ and $\frac{1}{4}$ respectively. The message is contaminated by a noise n, which is independent from m and takes on the values -1, 0, 1 with probabilities $\frac{1}{8}$, $\frac{5}{8}$, and $\frac{2}{8}$ respectively. The received signal, or the observation, can be represented as $r = m + n$. From r, we wish to infer what the transmitted message m was (estimated state), denoted using $\hat{m}$. $\hat{m}$ also takes values on 0 or 1. When $m = \hat{m}$, the detector correctly receives the original message, otherwise an error occurs.

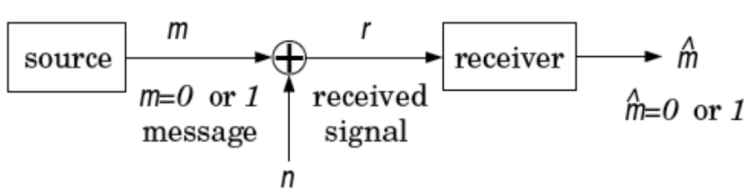

Figure 1: A simple receiver

a. Find the decision rule that achieves the maximum probability of correct decision. Compute the probability of error for this decision rule.

b. Let's have the noise n be a continuous random variable, uniformly distributed between $\frac{-3}{4}$ and $\frac{2}{4}$, and still statistically independent of m. First, plot the pdf of n. Then, find a decision rule that achieves the minimum probability of error, and compute the probability of error.

Problem 4: [16 points]

[Note: Use Matlab for the computations, but make sure to explicitly construct every transformation required, that is either type it or write it. Do not use Matlab 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

$$\mu = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}, \mathbf{\Sigma} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 4 \\ 0 & 4 & 5 \end{pmatrix}.$$

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

a. Find the probability density at the point $\mathbf{x_0} = (5\ 6\ 3)^T$

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. Remember that $\mathbf{A_w} = \mathbf{\Phi\Lambda^{-1/2}}$ is a linear transformation (i.e., $\mathbf{A_w}(a\mathbf{x} + b\mathbf{y}) = a\mathbf{A_w}\mathbf{x} + b\mathbf{A_w}\mathbf{y}$)

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?

Problem 5: [16 Points]

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

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

b. Use error functions $erf(*)$ to express d' in terms of the hit and false alarm rates. Estimate d' if $P(x > x^*|x \in w_1) = .7$ and $P(x > x^*|x \in w_2) = .5$. Repeat for d' if $P(x > x^*|x \in w_1) = .9$ and $P(x > x^*|x \in w_2) = .15$.

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:

$$\text{Case A: } P(x > x^*|x \in w_1) = .75 \text{ and } P(x > x^*|x \in w_2) = .35.$$

$$\text{Case B: } P(x > x^*|x \in w_1) = .8 \text{ and } P(x > x^*|x \in w_2) = .25.$$

Problem 6: [16 Points]

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 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.mat. This file contains n=5000 samples from a 1-dimensional Gaussian distribution.

Problem 7: [10 Points]

Suppose $\mathbf{x}$ and $\mathbf{y}$ are random variables. Their joint density, depicted below, is constant in the shaded area and 0 elsewhere,

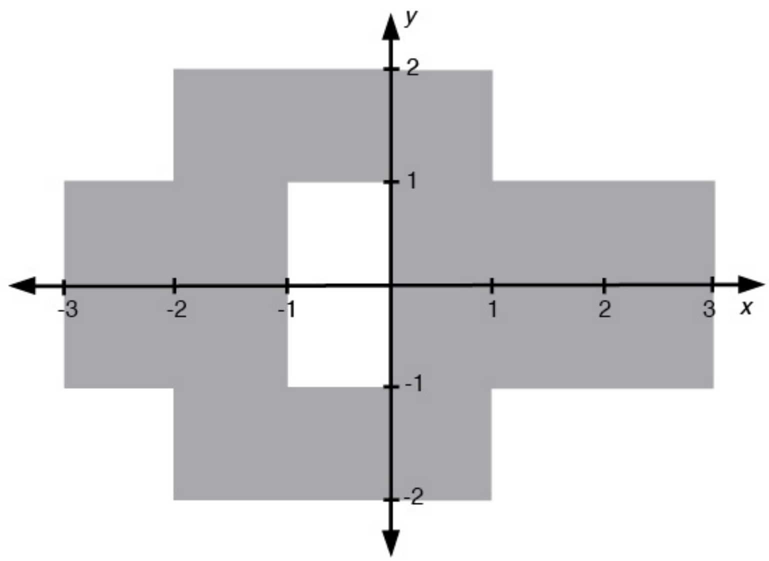

Figure 2: The joint distribution of $\mathbf{x}$ and $\mathbf{y}$.

a. Let ω_1 be the case when $\mathbf{x} \leq 0$, and ω_2 be the case when $\mathbf{x} > 0$. Determine the *a priori* probabilities of the two classes $\mathrm{P}(\omega_1)$ and $\mathrm{P}(\omega_2)$. Let $\mathbf{y}$ be the observation from which we infer whether ω_1 or ω_2 happens. Find the likelihood functions, namely, the two conditional distributions $p(y|\omega_1)$ and $p(y|\omega_2)$.

b. Find the decision rule that minimizes the probability of error, and calculate what the probability of error is. Please note that there will be ambiguities at decision boundaries, but how you classify when y falls on the decision boundary doesn't affect the probability of error.