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.

Solution: Average risk is choosing class ω_i:

$$R(\omega_i|\mathbf{x}) = \sum_{j=1}^{c} \lambda(\omega_i|\omega_j) P(\omega_j|\mathbf{x}) = 0^* P(\omega_i|\mathbf{x}) + \sum_{j=1, j!=i}^{c} \lambda_s P(\omega_j|\mathbf{x})$$

where $\lambda(\omega_i|\omega_j)$ is used to mean the cost of choosing class ω_i where the true class is ωj.

Hence:

$$R(\omega_i|\mathbf{x}) = \lambda_s(1 - P(\omega_i|\mathbf{x}))$$

Associate $\mathbf{x}$ with the class ω_i if highest posterior class probability and the average risk is less than the cost of rejection:

$$\lambda_s(1 - P(\omega_i|\mathbf{x})) \leq \lambda_r$$

$$P(\omega_i|\mathbf{x}) \geq 1 - \lambda_r/\lambda_s$$

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.

Solution: We solve for the parameters k_1 and k_2 by recognizing that the two equations are in the form of the normal Gaussian distribution.

$$\begin{aligned} k_1 e^{\frac{-\mathbf{x}^2}{20}} &= \frac{1}{\sqrt{2\pi {\sigma_1}^2}} e^{\frac{-(\mathbf{x}-\mu_1)^2}{2\sigma_1^2}} \\ \frac{-\mathbf{x}^2}{20} &= \frac{-(\mathbf{x}-\mu_1)^2}{2\sigma_1^2} \\ \mu_1 &= 0; \sigma_1^2 = 10 \\ k_1 &= \frac{1}{\sqrt{20\pi}} \end{aligned}$$

Figure 1: The pdf and decision boundaries

In a similar fashion, we find that

$$\mu_2 = 6; \sigma_2^2 = 6$$
$$k_2 = \frac{1}{\sqrt{12\pi}}$$

These distributions are plotted in Figure 1.

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?

Solution: The conditional risk for *Two-Category Classification* is discussed in Section 2.2.1 in D.H.S. where we can see that the formulas for the risk associated with the action, α_i, of classifying a feature vector, x, as class, ω_i, is different for each action:

$$R(\alpha_1|x) = \lambda_{11}P(\omega_1|x) + \lambda_{12}P(\omega_2|x)$$
$$R(\alpha_2|x) = \lambda_{21}P(\omega_1|x) + \lambda_{22}P(\omega_2|x)$$

The prior probabilities of the two classes are equal:

$$P(\omega_1) = P(\omega_2) = \frac{1}{2}$$

The assumption that the cost for choosing correctly is zero:

$$\lambda_{11} = 0, \lambda_{22} = 0$$

The costs for choosing incorrectly are given as C_{12} and C_{21}:

$$\lambda_{12} = C_{12} = \sqrt{3}$$
$$\lambda_{21} = C_{21} = \sqrt{5}$$

Thus the expression for the conditional risk of α_1 is:

$$\begin{aligned} R(\alpha_1|\mathbf{x}) &= \lambda_{11}P(\omega_1|\mathbf{x}) + \lambda_{12}P(\omega_2|\mathbf{x}) \\ R(\alpha_1|\mathbf{x}) &= 0P(\omega_1|\mathbf{x}) + \sqrt{3}P(\omega_2|\mathbf{x}) \\ R(\alpha_1|\mathbf{x}) &= \sqrt{3}P(\omega_2|\mathbf{x}) \end{aligned}$$

And the expression for the conditional risk of α_2:

$$\begin{aligned} R(\alpha_2|\mathbf{x}) &= \lambda_{21}P(\omega_1|\mathbf{x}) + \lambda_{22}P(\omega_2|\mathbf{x}) \\ R(\alpha_2|\mathbf{x}) &= \sqrt{5}P(\omega_1|\mathbf{x}) + 0P(\omega_2|\mathbf{x}) \\ R(\alpha_2|\mathbf{x}) &= \sqrt{5}P(\omega_1|\mathbf{x}) \end{aligned}$$

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

Solution: The Bayes Risk is the integral of the conditional risk when we use the optimal decision regions, R_1 and R_2. So, solving for the optimal decision boundary is a matter of solving for the roots of the equation:

$$\begin{aligned} R(\alpha_1|\mathbf{x}) &= R(\alpha_2|\mathbf{x}) \\ \sqrt{3}P(\omega_2|\mathbf{x}) &= \sqrt{5}P(\omega_1|\mathbf{x}) \\ \frac{\sqrt{3}P(\mathbf{x}|\omega_2)P(\omega_2)}{P(\mathbf{x})} &= \frac{\sqrt{5}P(\mathbf{x}|\omega_1)P(\omega_1)}{P(\mathbf{x})} \end{aligned}$$

Given the priors are equal this simplifies to:

$$\sqrt{3}P(\mathbf{x}|\omega_2) = \sqrt{5}P(\mathbf{x}|\omega_1)$$

Next, using the values, k_1 and k_2, from part (a), we have expressions for $p_{\mathbf{x}|\omega_1}$ and $p_{\mathbf{x}|\omega_2}$.

$$\begin{aligned}
\sqrt{3}\frac{1}{\sqrt{12\pi}}e^{\frac{-(\mathbf{x}-6)^2}{12}} &= \sqrt{5}\frac{1}{\sqrt{20\pi}}e^{\frac{-\mathbf{x}^2}{20}} \\
e^{\frac{-(\mathbf{x}-6)^2}{12}} &= e^{\frac{-\mathbf{x}^2}{20}} \\
\frac{-(\mathbf{x}-6)^2}{12} &= \frac{-\mathbf{x}^2}{20} \\
20(x^2 - 12x + 36) &= 12x^2 \\
8x^2 - 240x + 720 &= 0
\end{aligned}$$

The decision boundary is found by solving for the roots of this quadratic, $x_1 = 15 - 3\sqrt{15} = 3.81$ and $x_2 = 15 + 3\sqrt{15} = 26.62$

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

Solution: For $\mathbf{x} < 15 - 3\sqrt{15}$ the decision rule will choose ω_1
For $15 - 3\sqrt{15} < \mathbf{x} < 15 + 3\sqrt{15}$ the decision rule will choose ω_2
For $15 + 3\sqrt{15} < \mathbf{x}$ the decision rule will choose ω_1

Thus the decision region R_1 is $\mathbf{x} < 15 - 3\sqrt{15}$ and $\mathbf{x} > 15 + 3\sqrt{15}$, and the decision region R_2 is $15 - 3\sqrt{15} < \mathbf{x} < 15 + 3\sqrt{15}$.

$$
\begin{aligned}
Risk &= \int_{R_1} \lambda_{11} P(\omega_1|\mathbf{x}) + \lambda_{12} P(\omega_2|\mathbf{x}) \\
&+ \int_{R_2} \lambda_{21} P(\omega_1|y) + \lambda_{22} P(\omega_2|y) \\
&= \int_{R_1} \lambda_{12} P(y|\omega_2) P(\omega_2) + \int_{R_2} \lambda_{21} P(y|\omega_1) P(\omega_1) \\
&= \int_{y=-\inf}^{y=15-3\sqrt{15}} \sqrt{3} N(6,6) \frac{1}{2} + \int_{y=15+3\sqrt{15}}^{y=\inf} \sqrt{3} N(6,6) \frac{1}{2} \\
&+ \int_{y=15-3\sqrt{15}}^{y=15+3\sqrt{15}} \sqrt{5} N(0,10) \frac{1}{2} \\
&= \frac{\sqrt{3}}{2} \left((\frac{1}{2} + \frac{1}{2} erf(((15-3\sqrt{15})-6)/\sqrt{12})) + (\frac{1}{2} - \frac{1}{2} erf(((15+3\sqrt{15})-6)/\sqrt{12})) \right) \\
&+ \frac{\sqrt{5}}{2} \left(\frac{1}{2} erf((15+3\sqrt{15})/\sqrt{20}) - \frac{1}{2} erf((15-3\sqrt{15})/\sqrt{20}) \right) \\
&= 0.6824
\end{aligned}
$$

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.

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

   Solution: It is equivalent to find the decision rule that achieves the minimum probability of error. The receiver decides the transmitted message is 1, i.e., $\hat{m} = 1$ if

Figure 2: A simple receiver

$$\begin{aligned} P(r|m=1) \cdot Pr(m=1) &\geq P(r|m=0) \cdot Pr(m=0) \\ P(r|m=1) &\geq P(r|m=0) \cdot 3 \end{aligned} \tag{1}$$

Otherwise the receiver decides $\hat{m} = 0$. The likelihood functions for these two cases are

$$P(r|m=1) = \begin{cases} \frac{1}{8} & \text{if } r=0 \\ \frac{5}{8} & \text{if } r=1 \\ \frac{2}{8} & \text{if } r=2 \\ 0 & \text{otherwise} \end{cases}$$

$$P(r|m=0) = \begin{cases} \frac{1}{8} & \text{if } r=-1 \\ \frac{5}{8} & \text{if } r=0 \\ \frac{2}{8} & \text{if } r=1 \\ 0 & \text{otherwise} \end{cases}$$

Eq. 1 holds only when $r = 2$. Therefore, the decision rule can be summarized as

$$\hat{m} = \begin{cases} 1 & \text{if r} = 2 \\ 0 & \text{otherwise} \end{cases}$$

The probability of error is as follows:

$$\begin{aligned} Pr(e) &= Pr(\hat{m}=1|m=0) \cdot Pr(m=0) + Pr(\hat{m}=0|m=1) \cdot Pr(m=1) \\ &= Pr(r=2|m=0) \cdot Pr(m=0) + Pr(r \neq 2|m=1) \cdot Pr(m=1) \\ &= 0 + (\frac{1}{8} + \frac{5}{8}) \cdot \frac{1}{4} = \frac{3}{16} = 0.1875 \end{aligned}$$

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.

Solution:

The uniform distribution is plotted in Figure 3.

The decision rule is still determined by using Eq. 1. The likelihood functions become continuous, instead of discrete in (a):

Figure 3: The pdf of n

$$
\begin{aligned}
P(r|m=1) &= \begin{cases} \frac{4}{5} & \text{if } \frac{1}{4} < r \leq \frac{6}{4} \\ 0 & \text{otherwise} \end{cases} \\
P(r|m=0) &= \begin{cases} \frac{4}{5} & \text{if } \frac{-3}{4} < r \leq \frac{2}{4} \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

The interesting region is where the two pdf's overlap with each other, namely when $\frac{1}{4} < r \leq \frac{2}{4}$. From Eq. 1, we know we should decide $\hat{m} = 0$ for this range. The decision rule can be summarized as

$$
\hat{m} = \begin{cases} 0 & \text{if } \frac{-3}{4} < r \leq \frac{1}{4} \\ 0 & \text{if } \frac{1}{4} < r \leq \frac{2}{4} \\ 1 & \text{if } \frac{2}{4} < r \leq \frac{6}{4} \end{cases}
$$

Note that at the decision boundaries, there is ambiguity on which decision we should make. Again, either decision won't change the probability of error, so it is acceptable to decide both ways.

The probability of error is

$$\begin{aligned} Pr(e) &= Pr(\hat{m}=1|m=0)\cdot Pr(m=0) + Pr(\hat{m}=0|m=1)\cdot Pr(m=1) \\ &= Pr(\frac{1}{4} < r \leq \frac{2}{4}|m=1)\cdot Pr(m=1) \\ &= (\frac{2}{4}-\frac{1}{4})\cdot\frac{4}{5}\cdot\frac{1}{4} = \frac{1}{20} \end{aligned}$$

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:

Solution:

$$\mathbf{\Lambda} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 9 \end{pmatrix}, \mathbf{\Phi} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}.$$

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

Solution:

$$p(\mathbf{x}_0|\omega) = \frac{1}{(2\pi)^{3/2}|\Sigma|^{1/2}} e^{\frac{-(\mathbf{x}_0-\mu)^T\Sigma^{-1}(\mathbf{x}_0-\mu)}{2}} \tag{2}$$

$$|\Sigma| = 9;\ \Sigma^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{5}{9} & \frac{-4}{9} \\ 0 & \frac{-4}{9} & \frac{5}{9} \end{pmatrix}$$

The squared Mahalanobis distance from the mean to $\mathbf{x}_0$ is:

$$(\mathbf{x}_0 - \mu)^T \Sigma^{-1} (\mathbf{x}_0 - \mu) =$$
$$\left[\begin{pmatrix} 5 \\ 6 \\ 3 \end{pmatrix} - \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix} \right]^T \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{5}{9} & \frac{-4}{9} \\ 0 & \frac{-4}{9} & \frac{5}{9} \end{pmatrix} \left[\begin{pmatrix} 5 \\ 6 \\ 3 \end{pmatrix} - \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix} \right]$$
$$= 14$$

We plug these values in to find that the density at $\mathbf{x}_0$ is:

$$p(\mathbf{x}_0|\omega) = 1.2866e - 5$$

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

Solution:

$$\begin{aligned} y &= \Phi^T x \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}^T x \end{aligned}$$

To prove that for orthonormal transformations, Euclidean distances are preserved we have to show that $||y||^2 = ||x||^2$.

Proof: Let y be a random variable such that $y = A^T x$ where A^T is an orthonormal transformation (i.e., $A^T A = I$)

$$\begin{aligned} ||y||^2 &= y^T y \\ &= (A^T x)^T A^T x \\ &= x^T A A^T x \\ &= x^T x \\ &= ||x||^2 \end{aligned}$$

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., $\mathbf{A_w}(a\mathbf{x} + b\mathbf{y}) = a\mathbf{A_w x} + b\mathbf{A_w y}$)

Solution:

$$\begin{aligned} u &= \Lambda^{-1/2} y \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{\sqrt{9}} \end{pmatrix} y \end{aligned}$$

Where u can also be expressed as $u = \Lambda^{-1/2}\Phi^T x$

Let $u = \mathbf{A_w}^T x$ where $\mathbf{A_w} = \Phi\Lambda^{-1/2}$.

Given $p(x|w) \sim N(\mu, \Sigma)$ then $p(u|w) \sim N(\mathbf{A_w}^T\mu, \mathbf{A_w}^T\Sigma\mathbf{A_w})$

Since u is a result of a whitening transform to x, then the covariance matrix of u is proportional to the Identity matrix I. In this case, since Φ is a matrix of normalized eigenvectors of Σ and Λ is the eigenvalue matrix, the transformation $A^T = \Lambda^{-1/2}\Phi^T$ makes the covariance matrix equal to I.

Therefore, to convert the distribution to one centered on the origin with the covariance equal to the Identity matrix, it is enough to define a new variable z such that $z = \mathbf{A_w}^T x - \mathbf{A_w}^T\mu$. In that way $p(z|w) \sim N(0, 1)$.

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

Solution:

$$\begin{aligned} \mathbf{x_w} &= \mathbf{A_w}^T\mathbf{x_0} - \mathbf{A_w}^T\mu \\ &= \mathbf{A_w}^T(\mathbf{x_0} - \mu) \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{18}} & \frac{1}{\sqrt{18}} \end{pmatrix} \begin{pmatrix} 2 \\ 5 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} 2 \\ \frac{-4}{\sqrt{2}} \\ \frac{6}{\sqrt{18}} \end{pmatrix} \end{aligned}$$

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?

Solution: The squared Mahalanobis distance from $\mathbf{x_0}$ to the mean μ, we found in part (a), is .9514

The squared Mahalanobis distance from $\mathbf{x_w}$ to 0 is:

$$\mathbf{x_w}^T\mathbf{x_w} = \begin{pmatrix} 2 \\ \frac{-4}{\sqrt{2}} \\ \frac{6}{\sqrt{18}} \end{pmatrix}^T \begin{pmatrix} 2 \\ \frac{-4}{\sqrt{2}} \\ \frac{6}{\sqrt{18}} \end{pmatrix} = 14$$

The two distances are the same, as expected under any linear transformation.

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'

Solution: Let $x = x$ and $x^* = x^*$. Based on the Gaussian assumption, we see that $P(x > x^*|x \in w_i) = P\left(\frac{x-\mu_i}{\sigma_i} > \frac{x^*-\mu_i}{\sigma_i}|x \in w_i\right) \sim N(0,1)$ for $i = 1, 2$ Thus, we know $\left(\frac{x-\mu_2}{\sigma_2}\right)$ from the hit rate $P(x > x^*|x \in w_2)$ and $\left(\frac{x-\mu_1}{\sigma_1}\right)$ from the false alarm rate $P(x > x^*|x \in w_1)$, and these let us calculate the discriminability.

$$\begin{aligned} \left|\frac{x^*-\mu_1}{\sigma_1} - \frac{x^*-\mu_2}{\sigma_2}\right| &= \left|\frac{x^*-\mu_1}{\sigma} - \frac{x^*-\mu_2}{\sigma}\right| \\ &= \left|\frac{\mu_2-\mu_1}{\sigma}\right| \\ &= \frac{|\mu_2-\mu_1|}{\sigma} \\ &= d' \end{aligned}$$

Therefore d' is uniquely determined by the hit and false alarm rates.

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 $P(x > x^*|x \in w_2) = .15$.

Solution: There are a couple of ways in which you can proceed. This document will detail one of them.

Let y be a random variable such that $P(y > y^*|y \in w_i) \sim N(0,1)$ for $i = 1, 2$.

$$\begin{aligned}
\frac{1}{\sqrt{2\pi}} \int_0^{y^*} e^{-\frac{t^2}{2}} dt &= \frac{1}{\sqrt{\pi}} \int_0^{y^*} \frac{1}{\sqrt{2}} e^{-\frac{t^2}{2}} dt \\
&\text{let } u = \frac{t}{\sqrt{2}}; du = \frac{dt}{\sqrt{2}} \\
&= \frac{1}{\sqrt{\pi}} \int_0^{\frac{y^*}{\sqrt{2}}} e^{-u^2} du \\
&= \frac{1}{2} erf\left(\frac{y^*}{\sqrt{2}}\right)
\end{aligned}$$

Then, given that $P(x > x^*|x \in w_i) = P\left(\frac{x-\mu_i}{\sigma_i} > \frac{x^*-\mu_i}{\sigma_i}|x \in w_i\right) \sim N(0,1)$ for $i = 1, 2$.

$$
\begin{aligned}
P\left(\frac{x-\mu_i}{\sigma_i} > \frac{x^*-\mu_i}{\sigma_i}|x \in w_i\right) &= 1 - P\left(\frac{x-\mu_i}{\sigma_i} < \frac{x^*-\mu_i}{\sigma_i}|x \in w_i\right) \\
&= \begin{cases} 1 - \left(\frac{1}{2} + \frac{1}{2}erf\left(\frac{x^*-\mu_i}{\sqrt{2}\sigma_i}\right)\right) & \text{if } \frac{x^*-\mu_i}{\sigma_i} > 0 \\ 1 - \left(\frac{1}{2} - \frac{1}{2}erf\left(\frac{\mu_i-x^*}{\sqrt{2}\sigma_i}\right)\right) & \text{if } \frac{x^*-\mu_i}{\sigma_i} < 0 \end{cases} \\
&= \begin{cases} \frac{1}{2} - \frac{1}{2}erf\left(\frac{x^*-\mu_i}{\sqrt{2}\sigma_i}\right) & \text{if } \frac{x^*-\mu_i}{\sigma_i} > 0 \\ \frac{1}{2} + \frac{1}{2}erf\left(\frac{\mu_i-x^*}{\sqrt{2}\sigma_i}\right) & \text{if } \frac{x^*-\mu_i}{\sigma_i} < 0 \end{cases} \\
&= \begin{cases} \frac{1}{2}\left(1 - erf\left(\frac{x^*-\mu_i}{\sqrt{2}\sigma_i}\right)\right) & \text{if } \frac{x^*-\mu_i}{\sigma_i} > 0 \\ \frac{1}{2}\left(1 + erf\left(\frac{\mu_i-x^*}{\sqrt{2}\sigma_i}\right)\right) & \text{if } \frac{x^*-\mu_i}{\sigma_i} < 0 \end{cases}
\end{aligned}
$$

Therefore:

$$
\begin{aligned}
\frac{x^*-\mu_i}{\sigma_i} &= \sqrt{2}\left(erf^{-1}[1 - 2P(x > x^*|x \in w_i)]\right) \text{ if } P(x > x^*|x \in w_i) < .5 \\
\frac{\mu_i - x^*}{\sigma_i} &= \sqrt{2}\left(erf^{-1}[2P(x > x^*|x \in w_i) - 1]\right) \text{ if } P(x > x^*|x \in w_i) > .5
\end{aligned}
$$

Then to find the values of $erf^-1()$ use a table of erf, a table of the cumulative normal distribution, or use Matlab erf and $erfinv$ functions.

$$
\begin{aligned}
d_1^* &= \left|-\frac{\mu_1 - x^*}{\sigma} - \frac{x^* - \mu_2}{\sigma}\right| \\
&= \left|-\sqrt{2}\left(erf^{-1}[2(.7) - 1]\right) - \sqrt{2}\left(erf^{-1}[2(.5) - 1]\right)\right| \\
&= \left|-\sqrt{2}\left(erf^{-1}[.4]\right) - \sqrt{2}\left(erf^{-1}[0]\right)\right| \\
&= .5244
\end{aligned}
$$

$$
\begin{aligned}
d_2^* &= \left| -\frac{\mu_1 - x^*}{\sigma} - \frac{x^* - \mu_2}{\sigma} \right| \\
&= \left| -\sqrt{2}\left(erf^{-1}[2(.9) - 1]\right) - \sqrt{2}\left(erf^{-1}[1 - 2(.15)]\right) \right| \\
&= \left| -\sqrt{2}\left(erf^{-1}[.8]\right) - \sqrt{2}\left(erf^{-1}[.7]\right) \right| \\
&= 2.3180
\end{aligned}
$$

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

Solution: Given

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

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

Since the regions χ_1 and χ_2 are defined by our decision boundary x^*. We can see that

$$\epsilon_1 = \int_{\chi_2} p(x|w_1)dx = P(x < x^*|x \in w_1) \quad = 1 - P(x > x^*|x \in w_1)$$

$$\textit{Similarly}$$

$$\epsilon_2 = \int_{\chi_1} p(x|w_2)dx = P(x > x^*|x \in w_2)$$

Therefore,

$$\Pr[\text{error1}] = (1 - .7)P(w_1) + (.5)P(w_2)$$
$$\Pr[\text{error2}] = (1 - .9)P(w_1) + (.15)P(w_2)$$

And if the priors are equally likely,

$$\Pr[\text{error1}] = .4$$
$$\Pr[\text{error2}] = .125$$

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

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

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

Solution: From part (b) we can calculate the discriminability given the hit and false alarm rates, and see that Case B has a higher discriminability.

$$\begin{aligned} d_1^* &= \left| -\frac{\mu_1 - x^*}{\sigma} - \frac{x^* - \mu_2}{\sigma} \right| \\ &= \left| -\sqrt{2}\left(erf^{-1}[2(.75) - 1]\right) - \sqrt{2}\left(erf^{-1}[1 - 2(.35)]\right) \right| \\ &= \left| -\sqrt{2}\left(erf^{-1}[.5]\right) - \sqrt{2}\left(erf^{-1}[.3]\right) \right| \\ &= 1.0598 \end{aligned}$$

$$\begin{aligned} d_2^* &= \left| \frac{x^* - \mu_1}{\sigma} + \frac{\mu_2 - x^*}{\sigma} \right| \\ &= \left| \sqrt{2}\left(erf^{-1}[1 - 2(.25)]\right) + \sqrt{2}\left(erf^{-1}[2(.8) - 1]\right) \right| \\ &= \left| \sqrt{2}\left(erf^{-1}[.5]\right) + \sqrt{2}\left(erf^{-1}[.6]\right) \right| \\ &= 1.5161 \end{aligned}$$

Key point of this problem: The error function erf is related to the cumulative normal distribution. Assuming that the probability of x belonging to one of two classes is Gaussian, then knowing the values of the false alarm and hit rates for an arbitrary x^* is enough to compute the discriminability d'. Moreover, if the Gaussian assumption holds, a determination of the discriminability allows us to calculate the Bayes error rate.

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.

Solution: Sample mean is unbiased:

$$\begin{aligned} E[\hat{\mu}] &= E[\frac{1}{n}\sum_{i=1}^{n} x_i] \\ &= \frac{1}{n}\sum_{i=1}^{n} E[x_i] \\ &= \frac{1}{n}\sum_{i=1}^{n} \mu \\ &= \frac{1}{n} n\mu \\ &= \mu \end{aligned}$$

This document will show two ways to prove that the sample variance

is biased. Proof 1:

$$
\begin{aligned}
E[\sum_{i=1}^{n}(x_i - \hat{\mu})^2] &= E[\sum_{i=1}^{n}(x_i)^2] - nE[\hat{\mu}^2] \\
&= nE[(x_i)^2] - \frac{1}{n}E[(\sum_{i=1}^{n} x_i)^2] \\
&= n(var(x_i) + (E[x_i])^2) - \frac{1}{n}E[(\sum_{i=1}^{n} x_i)^2] \\
&= n\sigma^2 + \frac{1}{n}(nE[x_i])^2) - \frac{1}{n}E[(\sum_{i=1}^{n} x_i)^2] \\
&= n\sigma^2 - \frac{1}{n}[E[(\sum_{i=1}^{n} x_i)^2] - (E[\sum_{i=1}^{n} x_i])^2] \\
&= n\sigma^2 - \frac{1}{n}var(\sum_{i=1}^{n} x_i) \\
&= n\sigma^2 - \frac{1}{n}n\sigma^2 \\
&= (n-1)\sigma^2
\end{aligned}
$$

Proof 2:

$$
\begin{aligned}
E[\hat{\sigma_{ML}}^2] &= E[\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2] \\
&= \frac{1}{n}\sum_{i=1}^{n} E[x_i^2 - 2x_i\hat{\mu} + \hat{\mu}^2] \\
&= \frac{1}{n}\sum_{i=1}^{n} E[x_i^2] - 2\hat{\mu}E[\sum_{i=1}^{n} x_i] + \sum_{i=1}^{n} E[\hat{\mu}^2] \\
&= \frac{1}{n}[n(\sigma^2 + \mu^2) - 2nE[\hat{\mu}^2] + nE[\hat{\mu}^2]] \\
&= \frac{1}{n}[n\sigma^2 + n\mu^2 - nE[\hat{\mu}^2]] \\
&= \frac{1}{n}[n\sigma^2 + n\mu^2 - n(frac\sigma^2 n + \mu^2)] \\
&= \frac{n-1}{n}\sigma^2 \\
&Therefore \\
\text{unbiased: } \hat{\sigma}^2 &= \frac{n}{n-1}\hat{\sigma_{ML}}^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu})^2
\end{aligned}
$$

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.

Solution:

(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.

The dataset was generated from a Gaussian normal distribution $\sim N(9, 2)$. The following Matlab code calculates the ML estimate of the mean = 9, and produces Figure 6 indicating that the variance = 2.0.

```
load ps2.mat;
data=ps2;
n = length(data)
%ML estimate of mean:
MLmean = sum(data) / n
MLvar = [];
unMLvar = [];
%ML estimate of variance:
for(i=2:n)
    varML = 0;
    tempMean = sum(data(1:i)) / i;
    for (j=1:i)
        varML = varML + (data(j) - tempMean)^2;
    end

    MLvar(i) = varML/i;
    x = data(1:i);
    unMLvar(i) = cov(data(1:i));
end
figure
plot(1:1:n, MLvar(1:n), 'g', 1:1:n, unMLvar(1:n), 'r')
```

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,

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 $\mathrm{P}(\omega_2)$.

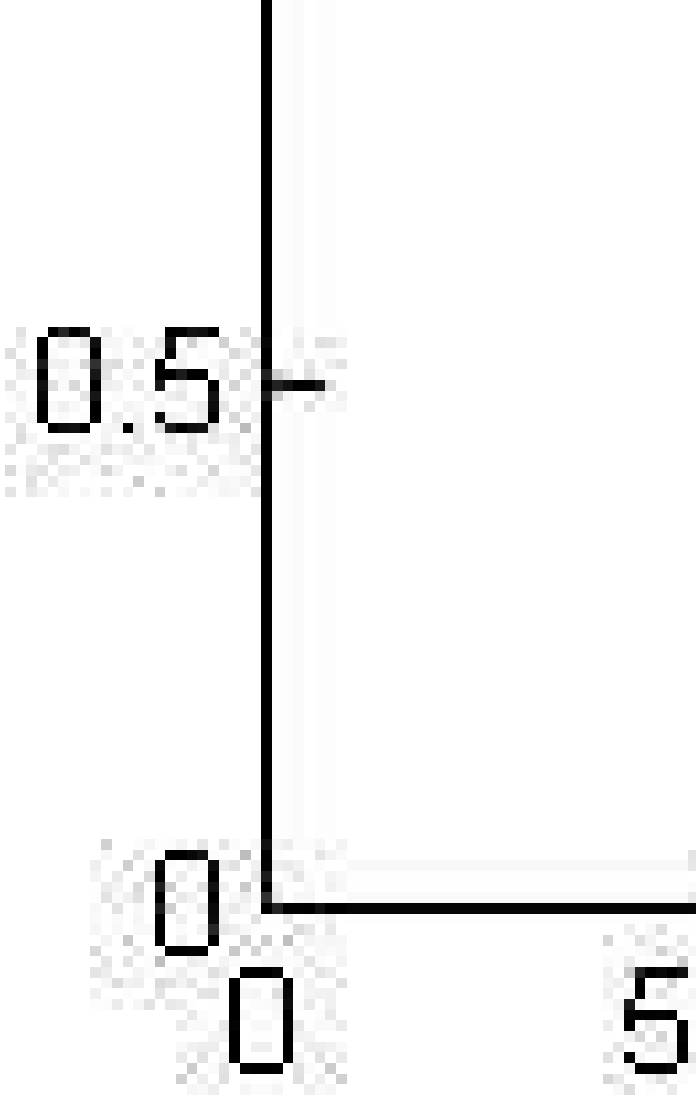

Figure 4: Though the two estimators have quite different estimates in the first few samples, the bias looses significance as n is large, and both estimators give the true variance of the data $\sigma = 1.0$.

Figure 5: The joint distribution of $\mathbf{x}$ and $\mathbf{y}$.

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 $p(y|\omega_2)$.

Solution:

By simply counting the number of unit squares in the shaded areas on the left and right sides of the line $x = 0$; we can directly find out that there are 8 unit squares on each side. Thus, the two cases are equally likely, i.e. $\mathrm{P}(\omega_1) = \mathrm{P}(\omega_2) = 0.5$.

It's also pretty straightforward to obtain the likelihood functions $p(y|\omega_1)$ and $p(y|\omega_2)$ by counting the number of unit squares for different ranges of y. We just need to be careful with normalizing them such that the integral of the distribution 1. See Figure 6.

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.

Solution: As shown in (a), the *a priori* probabilites of ω_1 and ω_2 are identical. The minimum probability of error decision rule simply relies on the comparison of the two likelihood function, In other words, it becomes a ML decision rule.

The decision rule can be summarized as:

$$\hat{\omega} = \begin{cases} \omega_1 \text{ if -2} < \text{y} \leq \text{-1 or 1} < \text{y} \leq 2 \\ \omega_2 \text{ if -1} < \text{y} \leq 1 \end{cases}$$

The probability of error is thus:

$$\begin{aligned} Pr(e) &= Pr(\hat{\omega} = \omega_1|\omega_2)Pr(\omega_1) + Pr(\omega_1|\omega_2)Pr(\omega_2) \\ &= \frac{1}{2}Pr(-1 < y \leq 1|\omega_1) + \frac{1}{2}Pr(-2 < y \leq -1, 1 < y \leq 2|\omega_2) \\ &= \frac{1}{2} * \frac{1}{2} + \frac{1}{2} * \frac{1}{4} = \frac{3}{8} \end{aligned}$$

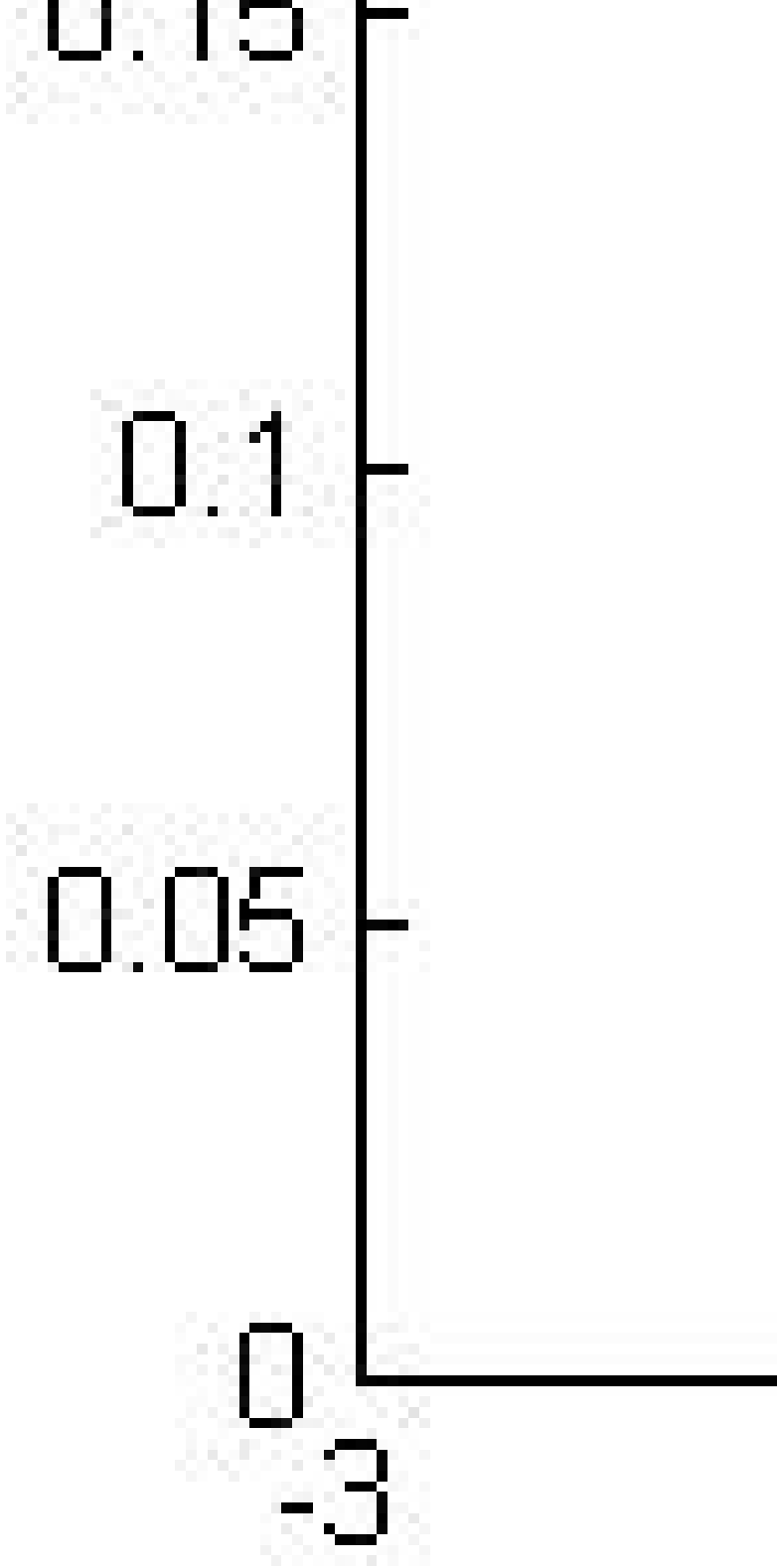

Figure 6: Likelihood functions: (a) $p(y|\ \omega_1)$ top, (b) $p(y|\ \omega_2)$ bottom