# 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 $\omega_{\ i}$, if $\mathrm{P}(\omega_{\ i}|\boldsymbol{x}) \geq \mathrm{P}(\omega_{\ j}|\boldsymbol{x})$ for all $j$ and and $\mathrm{P}(\omega_{\ i} — \boldsymbol{x}) \geq 1 - \frac{\lambda_r}{\lambda_s}$, and reject otherwise.

Solution:

Average risk is choosing class $\omega_i$:

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

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

Hence:

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

Associate $\mathbf{x}$ with the class $\omega_i$ 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]

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 $\mathbf{d'}$

   Let $\mathbf{x} = x$ and $\mathbf{x}^* = x^*$. 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 $\mathbf{d'}$ in terms of the hit and false alarm rates. Estimate $\mathbf{d'}$ if $\mathbf{P}(\mathbf{x} > \mathbf{x}^*|\mathbf{x} \in w_1) = .7$ and $\mathbf{P}(\mathbf{x} > \mathbf{x}^*|\mathbf{x} \in w_2) = .5$. Repeat for $\mathbf{d'}$ if $\mathbf{P}(\mathbf{x} > \mathbf{x}^*|\mathbf{x} \in w_1) = .85$ and $\mathbf{P}(\mathbf{x} > \mathbf{x}^*|\mathbf{x} \in w_2) = .15$.

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 $i = 1, 2$.

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

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.

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

$$
\begin{aligned}
d_2^* &= \left| -\frac{\mu_1 - x^*}{\sigma} - \frac{x^* - \mu_2}{\sigma} \right| \\
&= \left| -\sqrt{2} \left( erf^{-1}[2(.85) - 1] \right) - \sqrt{2} \left( erf^{-1}[1 - 2(.15)] \right) \right| \\
&= \left| -\sqrt{2} \left( erf^{-1}[.7] \right) - \sqrt{2} \left( erf^{-1}[.7] \right) \right| \\
&= 0
\end{aligned}
$$

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

According to Problem 3

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

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

Since the regions $\chi_1$ and $\chi_2$ are defined by our decision boundary $x^*$. We can see that

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

*Similarly*

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

Therefore,

$$
\begin{aligned}
\Pr[\text{error1}] &= (1 - .7)P(w_1) + (.5)P(w_2) \\
\Pr[\text{error2}] &= (1 - .85)P(w_1) + (.15)P(w_2)
\end{aligned}
$$

And if the priors are equally likely,

$$
\begin{aligned}
\Pr[\text{error1}] &= .5 \\
\Pr[\text{error2}] &= .15
\end{aligned}
$$

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

$$\textbf{Case A: } \mathbf{P}(\mathbf{x} > \mathbf{x}^*|\mathbf{x} \in w_1) = .75 \text{ and } \mathbf{P}(\mathbf{x} > \mathbf{x}^*|\mathbf{x} \in w_2) = .45.$$

$$\textbf{Case B: } \mathbf{P}(\mathbf{x} > \mathbf{x}^*|\mathbf{x} \in w_1) = .2 \text{ and } \mathbf{P}(\mathbf{x} > \mathbf{x}^*|\mathbf{x} \in w_2) = .9.$$

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

$$\begin{aligned}
d_1^* &= \left| -\frac{\mu_1 - x^*}{\sigma} - \frac{x^* - \mu_2}{\sigma} \right| \\
&= \left| -\sqrt{2}\left(erf^{-1}[2(.7) - 1]\right) - \sqrt{2}\left(erf^{-1}[1 - 2(.4)]\right) \right| \\
&= \left| -\sqrt{2}\left(erf^{-1}[.4]\right) - \sqrt{2}\left(erf^{-1}[.2]\right) \right| \\
&= .777
\end{aligned}$$

$$\begin{aligned}
d_2^* &= \left| \frac{x^* - \mu_1}{\sigma} + \frac{\mu_2 - x^*}{\sigma} \right| \\
&= \left| \sqrt{2}\left(erf^{-1}[1 - 2(.2)]\right) + \sqrt{2}\left(erf^{-1}[2(.8) - 1]\right) \right| \\
&= \left| \sqrt{2}\left(erf^{-1}[.4]\right) + \sqrt{2}\left(erf^{-1}[.6]\right) \right| \\
&= 1.366
\end{aligned}$$

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 3: [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.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 4: [16 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 $\omega_1$ be the case when $\mathbf{x} \leq 0$, and $\omega_2$ be the case when $\mathbf{x} > 0$. Determine the *a priori* probabilities of the two classes $P(\omega_1)$ and $P(\omega_2)$. $\mathbf{y}$ is the observation. from which we infer whether $\omega_1$ or $\omega_2$ 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.

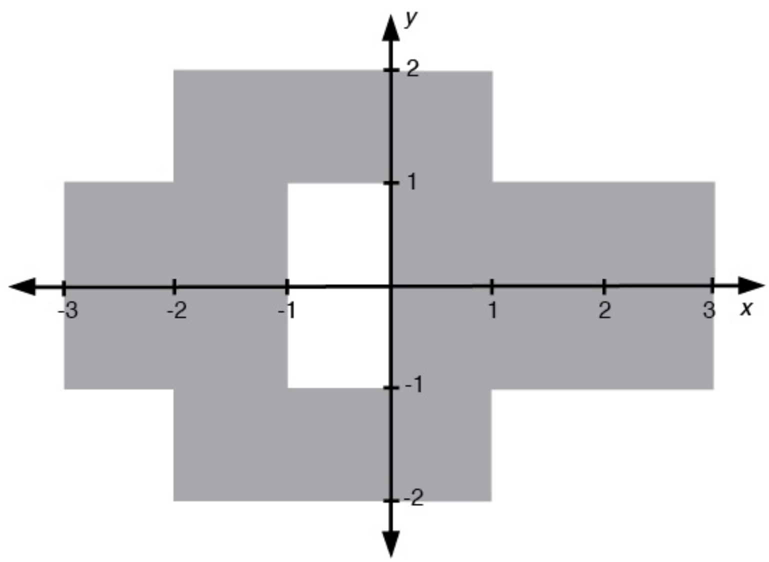


Figure 1: The joint distribution of $\mathbf{x}$ and $\mathbf{y}$.