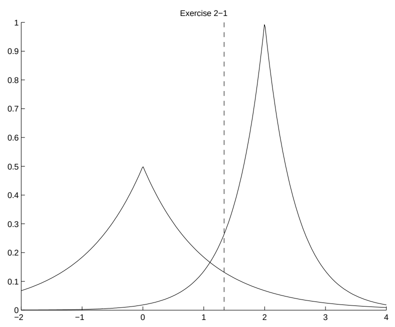
T-61.231 Principles of Pattern Recognition

Answers to exercise 2: 30.9.2002

1. **a)** Misclassification events are defined as follows:
   $\epsilon_1 = P(x \in \omega_1$, but is classified into class $\omega_2)$
   $\epsilon_2 = P(x \in \omega_2$, but is classified into class $\omega_1)$
   These probabilities are obtained by integration;
   $\epsilon_1 = \int_{R2} p(x|\omega_1)dx = \int_1^\infty 0.5e^{-|x|} = 0.5e^{-1}$
   $\epsilon_2 = \int_{R1} p(x|\omega_2)dx = \int_{-\infty}^1 e^{-2|x-2|} = \int_{-\infty}^1 0.5e^{2x-4} = 0.5e^{-2}$



   b) Let r be the decision boundary. From the figure it can be seen that when $r < 0 \Rightarrow \epsilon_1 > \frac{1}{2} > \epsilon_2$ and when $r > 2 \Rightarrow \epsilon_1 < \frac{1}{2} < \epsilon_2$ Therefore $r \in [0, 2]$ if $\epsilon_1 = \epsilon_2$
   $\epsilon_1 = \epsilon_2 \Leftrightarrow \int_r^\infty p(x|\omega_1)dx = \int_{-\infty}^r p(x|\omega_2)dx \Leftrightarrow 0.5e^{-r} = 0.5e^{2r-4} \Leftrightarrow r = \frac{4}{3}$

2. λ_{ij} is the cost of choosing class ω_j if the correct class is ω_i. The risks of both decisions are defined as
   $R(\text{Class } \omega_1 \text{ is chosen}) = \lambda_{11}p(\omega_1|x) + \lambda_{21}p(\omega_2|x) = R(\omega_1|x)$
   $R(\text{Class } \omega_2 \text{ is chosen}) = \lambda_{12}p(\omega_1|x) + \lambda_{22}p(\omega_2|x) = R(\omega_2|x)$

   According to the decision rule, the class corresponding to the lowest risk is chosen. Thus class ω_2 is chosen if $R(\omega_1|x) > R(\omega_2|x)$
   $\Leftrightarrow (\lambda_{11}-\lambda_{12})p(\omega_1|x) > (\lambda_{22}-\lambda_{21})p(\omega_2|x) \Leftrightarrow (\lambda_{11}-\lambda_{12})p(x|\omega_1)p(\omega_1) > (\lambda_{22}-\lambda_{21})p(x|\omega_2)p(\omega_2)$
   $\Leftrightarrow \frac{p(x|\omega_1)}{p(x|\omega_2)} < \frac{\lambda_{22}-\lambda_{21}}{\lambda_{11}-\lambda_{12}} \frac{p(\omega_2)}{p(\omega_1)}$
   where in the last stage it is assumed that $\lambda_{11} - \lambda_{12} < 0$, ie. the cost of choosing the correct class is assumed to be lower than choosing a wrong class (as should be expected).

   Now $p(x|\omega_1) = 0.5e^{-1|x|}$, $p(x|\omega_2) = e^{-2|x-2|}$, $\frac{\lambda_{22}-\lambda_{21}}{\lambda_{11}-\lambda_{12}} = \frac{1}{2}$ and the classes have some *a priori* probabilities and can be stated as $p(\omega_1) = p$ and $p(\omega_2) = 1 - p$. Thus $R(\omega_1|x) > R(\omega_2|x)$
   $\Leftrightarrow \frac{0.5e^{-1|x|}}{e^{-2|x-2|}} < \frac{1}{2}\frac{1-p}{p} \Leftrightarrow e^{-1|x|+2|x-2|} < \frac{1}{p} - 1$
   Now by taking the natural logarithm of both sides and dividing the situation into three cases we can see that

   $x < 0 : x = 4 - \ln(\frac{1}{p} - 1)$
   $0 \le x \le 2 : x = \frac{4}{3} - \frac{1}{3}\ln(\frac{1}{p} - 1)$, the decision boundary
   $x > 2 : x = 4 + \ln(\frac{1}{p} - 1)$

   R_2 is reduced to an empty area if $\frac{p(x|\omega_1)}{p(x|\omega_2)} > \frac{1}{2}\frac{1-p}{p}$, $\forall x$. As $p(x|\omega_2)$ decreases faster than $p(x|\omega_2)$, it is sufficient that $\frac{0.5e^{-1|x|}}{e^{-2|x-2|}} > \frac{1}{2}\frac{1-p}{p}$ is true when $p(x|\omega_2)$ is at it's maximum value of 1: $y = 2 \Rightarrow p(x|\omega_1) = 0.5e^{-2}$. Then
   $\frac{p(2|\omega_1)}{p(2|\omega_2)} > \frac{1}{2}\frac{1-p}{p} \Rightarrow 0.5e^{-2} > \frac{1}{2}(\frac{1}{p} - 1) \Rightarrow p > \frac{1}{e^{-2}+1}$

3. Class ω_1 is selected if

$$\frac{p(x|\omega_1)}{p(x|\omega_2)} < \frac{\lambda_{22}-\lambda_{21}}{\lambda_{11}-\lambda_{12}}\frac{p(\omega_2)}{p(\omega_1)},\ p(x|\omega_i) = \frac{1}{\sqrt{2\pi}\sigma_i}e^{\frac{-x^2}{2\sigma_i^2}}$$

$$\sigma_1^2 > \sigma_2^2 \Rightarrow x^2 > \frac{2}{\frac{1}{\sigma_2^2}-\frac{1}{\sigma_1^2}}\ln\left(\frac{\lambda_{22}-\lambda_{21}}{\lambda_{11}-\lambda_{12}}\frac{p(\omega_2)}{p(\omega_1)}\frac{\sigma_1}{\sigma_2}\right)$$

$$\Leftrightarrow |x| > \sqrt{\frac{2}{2-1}\ln\left(7\frac{1-0.7}{0.7}\sqrt{\frac{1}{0.5}}\right)} = \sqrt{2\ln(3\sqrt{2})} \approx 1.70$$

Thus the given samples are classified as $\{\hat{x_1} = 2.8, \hat{x_5} = -1.9\} \in \omega_1$ and $\{\hat{x_2} = 0.2, \hat{x_3} = 1.4, \hat{x_4} = -0.6\} \in \omega_2$.

4. **a)** A Gaussian distribution for a d-dimensional random variable $\overline{x}$ can be written as

$$p(\overline{x}) = (2\pi)^{-d/2}|\Sigma|^{-\frac{1}{2}}e^{-\frac{1}{2}(\overline{x}-\overline{\mu})^T\Sigma^{-1}(\overline{x}-\overline{\mu})}$$

The Bayes decision rule with $\frac{\lambda_{22}-\lambda_{21}}{\lambda_{11}-\lambda_{12}} = 1$ and $\frac{p(\omega_2)}{p(\omega_1)} = 1$ is

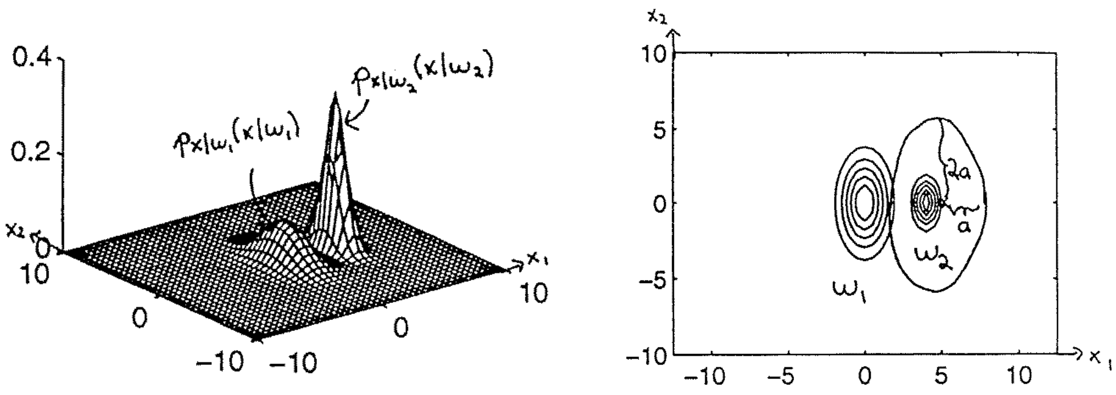
if $p(\overline{x}|\omega_1) > p(\overline{x}|\omega_2)$, select ω_1,
if $p(\overline{x}|\omega_1) < p(\overline{x}|\omega_2)$, select ω_2

On the decision surface $p(\overline{x}|\omega_1) = p(\overline{x}|\omega_2)$. And since we can also use any monotonically increasing function, lets use the natural logarithm, resulting in

$$\begin{aligned}
\ln(p(\overline{x}|\omega_1)) &= \ln(p(\overline{x}|\omega_2)) \\
-\tfrac{1}{2}(\overline{x}-\overline{\mu_1})^T\Sigma_1^{-1}(\overline{x}-\overline{\mu_1}) - \tfrac{1}{2}\ln|\Sigma_1| &= -\tfrac{1}{2}(\overline{x}-\overline{\mu_2})^T\Sigma_2^{-1}(\overline{x}-\overline{\mu_2}) - \tfrac{1}{2}\ln|\Sigma_2| \\
[x_1\ x_2]\begin{bmatrix}1 & 0\\ 0 & 4\end{bmatrix}^{-1}\begin{bmatrix}x_1\\ x_2\end{bmatrix} + \ln\begin{vmatrix}1 & 0\\ 0 & 4\end{vmatrix} &= \\
[(x_1-4)\ x_2]\begin{bmatrix}1/4 & 0\\ 0 & 1\end{bmatrix}^{-1}\begin{bmatrix}x_1-4\\ x_2\end{bmatrix} &+ \ln\begin{vmatrix}1/4 & 0\\ 0 & 1\end{vmatrix} \\
x_1^2 + \tfrac{1}{4}x_2^2 - 4x_1^2 + 16x_1 + 16x_1 + 64 - x_2^2 &= -2\ln 4 \\
(x_1 - \tfrac{16}{3})^2 + \tfrac{x_2^2}{4} &= \tfrac{1}{3}(2\ln 4 - 64) + (\tfrac{16}{3})^2 = a^2 \approx 8.03
\end{aligned}$$

Thus the decision surface is

$$\frac{(x_1-\frac{16}{3})^2}{a^2} + \frac{x_2^2}{(2a)^2} = 1$$



b) As in a),

$$
\begin{aligned}
\ln(p(\overline{x}|\omega_1)) &= \ln(p(\overline{x}|\omega_2)) \\
-\tfrac{1}{2}(\overline{x}-\overline{\mu_1})^T\Sigma_1^{-1}(\overline{x}-\overline{\mu_1}) - \tfrac{1}{2}\ln|\Sigma_1| &= -\tfrac{1}{2}(\overline{x}-\overline{\mu_2})^T\Sigma_2^{-1}(\overline{x}-\overline{\mu_2}) - \tfrac{1}{2}\ln|\Sigma_2| \\
[x_1\ (x_2-4)]\begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}^{-1}\begin{bmatrix} x_1 \\ x_2-4 \end{bmatrix} + \ln\begin{vmatrix} 1 & 0 \\ 0 & 4 \end{vmatrix} &= \\
& [(x_1-4)\ x_2]\begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}^{-1}\begin{bmatrix} x_1-4 \\ x_2 \end{bmatrix} + \ln\begin{vmatrix} 4 & 0 \\ 0 & 1 \end{vmatrix} \\
x_1^2 + \tfrac{1}{4}(x_2-4)^2 &= \tfrac{1}{4}(x_1-4)^2 + x_2 \\
(x_1+\tfrac{4}{3})^2 &= (x_2+\tfrac{4}{3})^2
\end{aligned}
$$

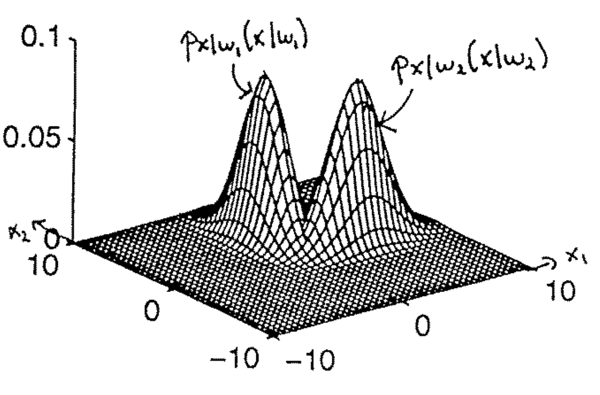
Thus the decision surface is

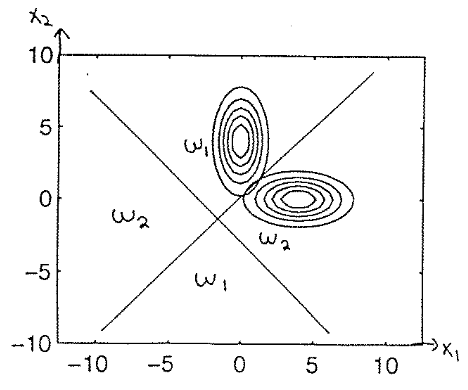
$$
\begin{cases} x_1 + \frac{4}{3} = x_2 + \frac{4}{3} \\ x_1 + \frac{4}{3} = -(x_2 + \frac{4}{3}) \end{cases} \Rightarrow \begin{cases} x_2 = x_1 \\ x_2 = -x_1 - \frac{8}{3} \end{cases}
$$

As the two lines divide the space into four segments, which class each segment belongs to can easily be verified by calculating one point in each segment and checking which class that point would be assigned to. Let's choose the points $(0, -\frac{4}{3})$, $(-\frac{4}{3}, 0)$, $(-\frac{8}{3}, -\frac{4}{3})$ and $(-\frac{4}{3}, -\frac{8}{3})$.

	x_1	x_2	$(x_1+\frac{4}{3})^2$	$(x_2+\frac{4}{3})^2$	ω
I	0	$-\frac{4}{3}$	$\frac{16}{9}$	0	ω_2
I	$-\frac{4}{3}$	0	0	$\frac{16}{9}$	ω_1
I	$-\frac{8}{3}$	$-\frac{4}{3}$	$\frac{16}{9}$	0	ω_2
I	$-\frac{4}{3}$	$-\frac{8}{3}$	0	$\frac{16}{9}$	ω_1

(If $(x_1+\frac{4}{3})^2 > (x_2+\frac{4}{3})^2$, $x \in \omega_2$. If $(x_1+\frac{4}{3})^2 < (x_2+\frac{4}{3})^2$, $x \in \omega_1$.)





5. The Poisson distribution: $p(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$, $x \geq 0$. The maximum likelihood estimate for the parameter λ: $p(H|\lambda) = \prod_{k=1}^{n} p(x_k|\lambda)$, where $H = \{x_1, \ldots, x_n\}$ is the sample set.

   The maximum likelihood estimate can be found by maximizing $p(H|\lambda)$. The same solution can also be found by maximizing the functions natural logarithm, as the logarithm function is monotonically increasing. The maximum can be found by setting the derivative regarding λ to zero,

   $\frac{\delta}{\delta\lambda} \ln\{p(H|\lambda)\} = \frac{\delta}{\delta\lambda} \sum_{k=1}^{n} (x_k \ln(\lambda) - \lambda - \ln(x_k!)) = \sum_{k=1}^{n} (x_k \frac{1}{\lambda} - 1) = \frac{1}{\lambda} \sum_{k=1}^{n} x_k - n = 0$
   $\Rightarrow \hat{\lambda} = \frac{1}{n} \sum_{k=1}^{n} x_k$

   For an unbiased estimate $E\{\hat{\theta}\} = \theta$.

   $E\{\hat{\lambda}\} = E\{\frac{1}{n} \sum_{k=1}^{n} x_k\} = \frac{1}{n} \sum_{k=1}^{n} E\{x_k\} = E\{x_k\} = \lambda$, as for the Poisson distribution $E\{x\} = \lambda$.

   Thus $\hat{\lambda}$ is unbiased.