University of Stavanger

Stavanger, December 17, 2024

Solutions to theoretical exercise 1

ELE520 Machine learning

Problem 1

a) We have the two classes: ω_1 = "toxic mussels" and ω_2 = "non toxic mussels". α_i corresponds to deciding ω_i

The loss functions are:

- Loss of rejecting non toxic container: $\lambda(\alpha_1|\omega_2) = 250$.
- Loss of poisoning customer: $\lambda(\alpha_2|\omega_1) = 100000$.
- Loss of correct classification: $\lambda(\alpha_1|\omega_1) = \lambda(\alpha_2|\omega_2) = 0$.

b) Classifies to ω_i, depending on measured x, toxic concentration. The expression for the conditional risk associated with deciding ω_1 when x is measured:

$$\begin{aligned} R(\alpha_1|x) &= \sum_{j=1}^{c} \lambda(\alpha_1|\omega_j) \cdot P(\omega_j|x) \qquad (1) \\ &= \lambda(\alpha_1|\omega_1) \cdot P(\omega_1|x) + \lambda(\alpha_1|\omega_2)P(\omega_2|x) \\ &= 0 + 250 \cdot P(\omega_2|x) \\ &= 250 \cdot \frac{p(x|\omega_2) \cdot P(\omega_2)}{p(x)} \\ &= 250 \cdot \frac{p(x|\omega_2) \cdot P(\omega_2)}{P(\omega_1) \cdot p(x|\omega_1) + P(\omega_2) \cdot p(x|\omega_2)} \end{aligned}$$

where $P(\omega_1) = \frac{1}{25}$ og $P(\omega_2) = \frac{24}{25}$. Substituted into the expression we get:

$$\begin{aligned} R(\alpha_1|x) &= \frac{250 \cdot \frac{1}{\sqrt{2\pi} \cdot 0.01} \cdot e^{-\frac{(x-0.2)^2}{0.0002}} \cdot \frac{24}{25}}{\frac{1}{25} \cdot \frac{1}{\sqrt{2\pi} \cdot 0.01} \cdot e^{-\frac{(x-0.4)^2}{0.0002}} + \frac{24}{25} \cdot \frac{1}{\sqrt{2\pi} \cdot 0.01} \cdot e^{-\frac{(x-0.2)^2}{0.0002}}} \\ &= \underline{\underline{\frac{6000 \cdot e^{-\frac{(x-0.2)^2}{0.0002}}}{e^{-\frac{(x-0.4)^2}{0.0002}} + 24 \cdot e^{-\frac{(x-0.2)^2}{0.0002}}}}} \end{aligned}$$

Correspondingly, we have the conditional risk associated with deciding ω_2 given

Trygve Eftestøl, Professor
Faculty of Science and Technology
Department of Electrical and Computer Engineering

Kjølv Egelands hus

University of Stavanger
N-4036 Stavanger.

Telephone: +47 51 83 10 00.
Telefax: +47 51 83 17 50.
E-mail: trygve.eftestol@uis.no
www.ux.his.no/~trygve-e

measurement x:

$$\begin{aligned}
R(\alpha_2|x) &= \lambda(\alpha_2|\omega_1)\cdot P(\omega_1|x) + \lambda(\alpha_2|\omega_2)P(\omega_2|x)\\
&= 100000\cdot\frac{p(x|\omega_1)\cdot P(\omega_1)}{p(x)}\\
&= \underline{\underline{\frac{100000\cdot e^{-\frac{(x-0.4)^2}{0.0002}}}{e^{-\frac{(x-0.4)^2}{0.0002}} + 24\cdot e^{-\frac{(x-0.2)^2}{0.0002}}}}}
\end{aligned}$$

c) The decision border minimising the average risk:

$$\begin{aligned}
R(\alpha_1|x) &= R(\alpha_2|x)\\
6000\cdot e^{-\frac{(x-0.2)^2}{0.0002}} &= 100000\cdot e^{-\frac{(x-0.4)^2}{0.0002}}\\
\ln\left(6000\cdot e^{-\frac{(x-0.2)^2}{0.0002}}\right) &= \ln\left(100000\cdot e^{-\frac{(x-0.4)^2}{0.0002}}\right)\\
\ln(6000) - \frac{(x-0.2)^2}{0.0002} &= \ln(100000) - \frac{(x-0.4)^2}{0.0002}\\
(x-0.4)^2 - (x-0.2)^2 &= (\ln(100000) - \ln(6000))\cdot 0.0002\\
x^2 - 0.8x + 0.16 - x^2 + 0.4x - 0.04 &= 2.8134\cdot 0.0002\\
x^2 - 0.8x + 0.16 - x^2 + 0.4x - 0.04 - 0.0006 &= 0\\
-0.4x + 0.1194 &= 0\\
x = x_0 &= \underline{0.2986}
\end{aligned}$$

Thus we assume the shells are non toxic for $x < 0.2986$, and toxic shells otherwise.

d) The minimum average cost R associated with classification (in NOK):

$$\begin{aligned}
R &= \int R(\alpha(x)|x)\cdot p(x)dx \qquad (2)\\
&= \int_{-\infty}^{x_0} R(\alpha_2|x)\cdot p(x)dx + \int_{x_0}^{\infty} R(\alpha_1|x)\cdot p(x)dx\\
&= \frac{1}{25}100000\cdot\int_{-\infty}^{x_0} p(x|\omega_1)dx + \frac{24}{25}250\cdot\int_{x_0}^{\infty} p(x|\omega_2)dx\\
&= 4000\cdot P(x < x_0|\omega_1) + 240\cdot P(x > x_0|\omega_2)\\
&= 4000\cdot P(x < x_0|\omega_1) + 240\cdot(1 - P(x < x_0|\omega_2))\\
&= 4000\cdot P\left(\frac{x-0.4}{0.01} < \frac{x_0-0.4}{0.01}\right)\\
&\quad + 240\cdot\left(1 - P\left(\frac{x-0.2}{0.01} < \frac{x_0-0.2}{0.01}\right)\right)\\
&= 4000\cdot P\left(y < \frac{x_0-0.4}{0.01}\right) + 240\cdot\left(1 - P\left(y < \frac{x_0-0.2}{0.01}\right)\right)\\
&= 4000\cdot P(y < -10.14) + 240\cdot(1 - P(y < 9.86))\\
&= 4000\cdot(1 - P(y < 10.14)) + 240\cdot(1 - P(y < 9.86))\\
&\approx 4000\cdot(1-1) + 240\cdot(1-1) = \underline{\underline{0}}
\end{aligned}$$

Here $y \sim N(0,1)$. (The value is so small that it does not appear in the table for the normal distribution.) The results shows that the classification system is practically without risk. (There is nearly no overlap bertween the class conditional distributions).

Problem 2

a) The contour lines of the two class conditional probability density functions are shown in figure 3. The eigenvectors scaled by their respective eigenvalues have their origin in the mean values and end points in the coordinates $\boldsymbol{p}_i = \boldsymbol{\mu} + \sqrt{\lambda_i}\mathbf{e}_i, i = 1, 2$ which are $(3 + \sqrt{1/2}\ 3)^T$ and $(3\ 3 + \sqrt{2})^T$for class ω_1 and $(3 + \sqrt{2}\ \ -2)^T$ and $(3\ \ -2 + \sqrt{2})^T$for class ω_2.

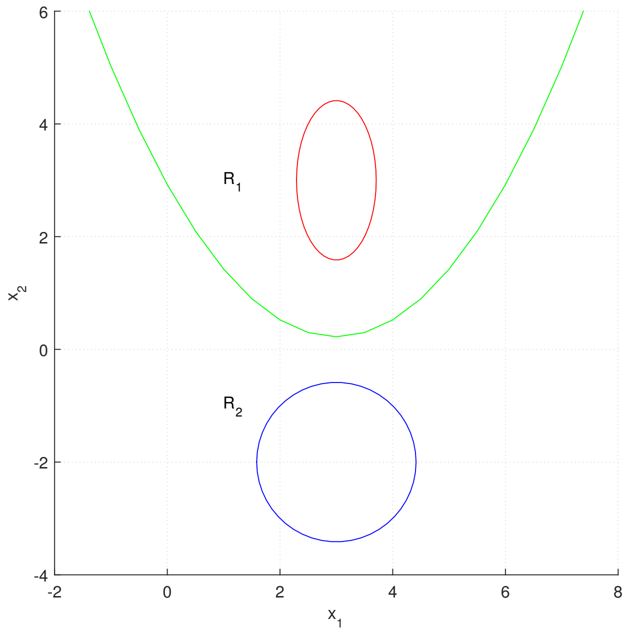

Figure 1: Contour lines for the class conditional density functions.

b) Bayes decision rule can be formulated as

$$\text{Decide} \begin{cases} \omega_1, & \text{if } P(\omega_1|\boldsymbol{x}) > P(\omega_2|\boldsymbol{x}) \\ \omega_2, & \text{otherwise} \end{cases} \tag{1}$$

The decision border can be found by solving the equation $P(\omega_1|\boldsymbol{x}) = P(\omega_2|\boldsymbol{x})$. We will rather use the discriminant functions for equal covariance matrices and solve the equation $g_1(\boldsymbol{x}) = g_2(\boldsymbol{x})$ where

$$\begin{aligned}
g_i(\boldsymbol{x}) &= -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| + \ln P(\omega_i) \\
&= \boldsymbol{x}^T\boldsymbol{\Theta}_i\boldsymbol{x} + \boldsymbol{\theta}_i{}^T\boldsymbol{x} + w_{i0} \\
&\text{der} \\
\boldsymbol{\Theta}_i &= -\frac{1}{2}\boldsymbol{\Sigma}_i^{-1} \\
\boldsymbol{\theta}_i &= \boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i \\
\theta_{i0} &= -\frac{1}{2}\boldsymbol{\mu}_i{}^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| + \ln P(\omega_i)
\end{aligned} \tag{2}$$

We compute

$$\begin{aligned}
\boldsymbol{\Theta}_1 &= -\frac{1}{2}\boldsymbol{\Sigma}_1^{-1} \\
&= -\frac{1}{2}\begin{pmatrix} 2 & 0 \\ 0 & 0.5 \end{pmatrix} \\
&= \begin{pmatrix} -1 & 0 \\ 0 & -0.25 \end{pmatrix},
\end{aligned} \tag{3}$$

$$\begin{aligned}
\boldsymbol{\theta}_1 &= \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 \\
&= \begin{pmatrix} 2 & 0 \\ 0 & 0.5 \end{pmatrix}\begin{pmatrix} 3 \\ 3 \end{pmatrix} \\
&= \begin{pmatrix} 6 \\ 1.5 \end{pmatrix}
\end{aligned} \tag{4}$$

and

$$\begin{aligned}
\theta_{10} &= -\frac{1}{2}\boldsymbol{\mu}_1{}^T\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| + \ln P(\omega_1) \\
&= -\frac{1}{2}\begin{pmatrix} 3 & 3 \end{pmatrix}\begin{pmatrix} 2 & 0 \\ 0 & 0.5 \end{pmatrix}\begin{pmatrix} 3 \\ 3 \end{pmatrix} - 0 - \ln 1/2 \\
&= -11.943
\end{aligned} \tag{5}$$

which gives

$$\begin{aligned}
g_1(\boldsymbol{x}) &= \boldsymbol{x}^T\boldsymbol{\Theta}_1\boldsymbol{x} + \boldsymbol{\theta}_1{}^T\boldsymbol{x} + \theta_{10} \\
&= \begin{pmatrix} x_1 & x_2 \end{pmatrix}\begin{pmatrix} -1 & 0 \\ 0 & -0.25 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 6 & 1.5 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - 11.943 \\
&= -x_1^2 - 0.25x_2^2 + 6x_1 + 1.5x_2 - 11.943
\end{aligned} \tag{6}$$

for class ω_1.

Compute

$$\begin{aligned}
\boldsymbol{\Theta}_2 &= -\frac{1}{2}\boldsymbol{\Sigma}_2^{-1} \\
&= -\frac{1}{2}\begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix} \\
&= \begin{pmatrix} -1/4 & 0 \\ 0 & -1/4 \end{pmatrix},
\end{aligned} \tag{7}$$

$$\begin{aligned}
\boldsymbol{\theta}_2 &= \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2 \\
&= \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}\begin{pmatrix} 3 \\ -2 \end{pmatrix} \\
&= \begin{pmatrix} 3/2 \\ -1 \end{pmatrix}
\end{aligned} \tag{8}$$

and

$$\begin{aligned}
\theta_{20} &= -\frac{1}{2}\boldsymbol{\mu}_2{}^T\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2 - \frac{1}{2}\ln|\boldsymbol{\Sigma}_2| + \ln P(\omega_2) \\
&= -\frac{1}{2}\begin{pmatrix} 3 & -2 \end{pmatrix}\begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}\begin{pmatrix} 3 \\ -2 \end{pmatrix} - 0 - \ln 1/2 \\
&= \begin{pmatrix} -3/2 & 1 \end{pmatrix}\begin{pmatrix} 3/2 \\ -1 \end{pmatrix} - 1/2\ln 4 - \ln 1/2 \\
&= -13/4 - 1/2\ln 4 - \ln 1/2 \\
&= -4.636
\end{aligned} \tag{9}$$

which gives

$$\begin{aligned}
g_2(\boldsymbol{x}) &= \boldsymbol{x}^T\boldsymbol{\Theta}_2\boldsymbol{x} + \boldsymbol{\theta}_2{}^T\boldsymbol{x} + \theta_{20} \\
&= \begin{pmatrix} x_1 & x_2 \end{pmatrix}\begin{pmatrix} -1/4 & 0 \\ 0 & -1/4 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} -2 & 3 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - 4.636 \\
&= -0.25x_1^2 - 0.25x_2^2 + 1.5x_1 - x_2 - 4.636
\end{aligned} \tag{10}$$

for class ω_2.

We then find the decision border by solving

$$\begin{aligned}
g_1(\boldsymbol{x}) &= g_2(\boldsymbol{x}) \\
-x_1^2 - 0.25x_2^2 + 6x_1 + 1.5x_2 - 11.942 &= -0.25x_1^2 - 0.25x_2^2 + 1.5x_1 - x_2 - 4.636 \\
&\Downarrow \\
x_2 &= 0.3x_1^2 - 1.8x_1 + 2.92.
\end{aligned} \tag{11}$$

The decision border is shown in figure 3. The decision regions, R_1, for class ω_1 and R_2, for class ω_2 are found to be above and below the decision border.

Problem 3

$$
\begin{aligned}
g_i'(\boldsymbol{x}) &= -\frac{\|\boldsymbol{x}-\boldsymbol{\mu}_i\|^2}{2\sigma^2} + \ln P(\omega_i) \\
&= -\frac{1}{2\sigma^2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^T(\boldsymbol{x}-\boldsymbol{\mu}_i) + \ln P(\omega_i) \\
&= -\frac{1}{2\sigma^2}[\boldsymbol{x}^T\boldsymbol{x} - 2\boldsymbol{\mu}_i{}^T\boldsymbol{x} + \boldsymbol{\mu}_i{}^T\boldsymbol{\mu}_i] + \ln P(\omega_i) \\
&\quad \text{remove class independent terms and get the equivalent discriminant function} \\
g_i(\boldsymbol{x}) &= -\frac{1}{2\sigma^2}[-2\boldsymbol{\mu}_i{}^T\boldsymbol{x} + \boldsymbol{\mu}_i{}^T\boldsymbol{\mu}_i] + \ln P(\omega_i) \\
&= \frac{1}{\sigma^2}\boldsymbol{\mu}_i{}^T\boldsymbol{x} + \frac{-1}{2\sigma^2}\boldsymbol{\mu}_i{}^T\boldsymbol{\mu}_i + \ln P(\omega_i)
\end{aligned} \tag{1}
$$

Furthermore we set $g_i(\boldsymbol{x}) = g_j(\boldsymbol{x})$ so that

$$
\frac{1}{\sigma^2}\boldsymbol{\mu}_i{}^T\boldsymbol{x} + \frac{-1}{2\sigma^2}\boldsymbol{\mu}_i{}^T\boldsymbol{\mu}_i + \ln P(\omega_i) = \frac{1}{\sigma^2}\boldsymbol{\mu}_j{}^T\boldsymbol{x} + \frac{-1}{2\sigma^2}\boldsymbol{\mu}_j{}^T\boldsymbol{\mu}_j + \ln P(\omega_j) \tag{2}
$$

corresponding to

$$
\begin{aligned}
\frac{1}{\sigma^2}(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)^T\boldsymbol{x} - \frac{1}{2\sigma^2}\boldsymbol{\mu}_i{}^T\boldsymbol{\mu}_i + \frac{1}{2\sigma^2}\boldsymbol{\mu}_j{}^T\boldsymbol{\mu}_j + \ln P(\omega_i) - \ln P(\omega_j) &= 0 \\
\text{which after multiplication with } \sigma^2 \text{ becomes} & \\
(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)^T\boldsymbol{x} - \frac{1}{2}\boldsymbol{\mu}_i{}^T\boldsymbol{\mu}_i + \frac{1}{2}\boldsymbol{\mu}_j{}^T\boldsymbol{\mu}_j + \sigma^2\ln\frac{P(\omega_i)}{P(\omega_j)} &= 0 \\
\text{adderer inn } \tfrac{1}{2}(\boldsymbol{\mu}_i{}^T\boldsymbol{\mu}_j - \boldsymbol{\mu}_j{}^T\boldsymbol{\mu}_i) \text{ which equals 0} & \\
(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)^T\boldsymbol{x} - \frac{1}{2}(\boldsymbol{\mu}_i{}^T\boldsymbol{\mu}_i - \boldsymbol{\mu}_j{}^T\boldsymbol{\mu}_j + \boldsymbol{\mu}_i{}^T\boldsymbol{\mu}_j - \boldsymbol{\mu}_j{}^T\boldsymbol{\mu}_i) + \sigma^2\ln\frac{P(\omega_i)}{P(\omega_j)} &= 0 \\
(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)^T\boldsymbol{x} - \frac{1}{2}(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)^T(\boldsymbol{\mu}_i+\boldsymbol{\mu}_j) + \sigma^2\ln\frac{P(\omega_i)}{P(\omega_j)} &= 0 \\
\text{manipulate the last term} & \\
(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)^T\boldsymbol{x} - \frac{1}{2}(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)^T(\boldsymbol{\mu}_i+\boldsymbol{\mu}_j) + \sigma^2\ln\frac{P(\omega_i)}{P(\omega_j)}\frac{(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)^T(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)}{(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)^T(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)} &= 0 \\
\text{and finally we get} & \\
(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)^T[\boldsymbol{x} - \{\frac{1}{2}(\boldsymbol{\mu}_i+\boldsymbol{\mu}_j) - \frac{\sigma^2}{\|\boldsymbol{\mu}_i-\boldsymbol{\mu}_j\|^2}\ln\frac{P(\omega_i)}{P(\omega_j)}(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j)\}] &= 0
\end{aligned} \tag{3}
$$

Problem 4

a)

$$
\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(\mathbf{x})d\mathbf{x} &= 1\\
\int_{a_2}^{b_2}\int_{a_1}^{b_1} c dx_1 dx_2 &= 1\\
c\int_{a_2}^{b_2} [x_1]_{a_1}^{b_1} dx_2 &= 1\\
c\int_{a_2}^{b_2} (a_1-b_1) dx_2 &= 1\\
c(b_1-a_1)[x_2]_{a_2}^{b_2} &= 1\\
c(b_1-a_1)(b_2-a_2) &= 1
\end{aligned}
\tag{1}
$$

i.e.

$$
c = \frac{1}{(b_1-a_1)(b_2-a_2)}.
$$

b)

$$
\begin{aligned}
\boldsymbol{\mu} &= \mathbb{E}[\boldsymbol{x}] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \boldsymbol{x}p(\boldsymbol{x})d\boldsymbol{x}\\
&= \int_{a_2}^{b_2}\int_{a_1}^{b_1} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \frac{1}{(b_1-a_1)(b_2-a_2)} dx_1 dx_2\\
&= \begin{bmatrix} \frac{1}{(b_1-a_1)(b_2-a_2)}[\frac{x_1^2}{2}]_{a_1}^{b_1}[x_2]_{a_2}^{b_2} \\ \frac{1}{(b_1-a_1)(b_2-a_2)}[\frac{x_2^2}{2}]_{a_2}^{b_2}[x_1]_{a_1}^{b_1} \end{bmatrix}\\
&= \begin{bmatrix} \frac{a_1+b_1}{2} \\ \frac{a_2+b_2}{2} \end{bmatrix}.
\end{aligned}
$$

This is the center of the rectangle where $p(\boldsymbol{x}) \neq 0$ as illustrated in figure 2.

c)

$$
\begin{aligned}
\boldsymbol{\Sigma} &= \mathbb{E}[(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T] = E\begin{bmatrix} (x_1-\mu_1)^2 & (x_1-\mu_1)(x_2-\mu_2) \\ (x_1-\mu_1)(x_2-\mu_2) & (x_2-\mu_2)^2 \end{bmatrix}\\
&= \begin{bmatrix} \mathbb{E}[x_1^2] & \mathbb{E}[x_1x_2] \\ \mathbb{E}[x_1x_2] & \mathbb{E}[x_2^2] \end{bmatrix} - \begin{bmatrix} \mu_1^2 & \mu_1\mu_2 \\ \mu_1\mu_2 & \mu_2^2 \end{bmatrix},
\end{aligned}
\tag{2}
$$

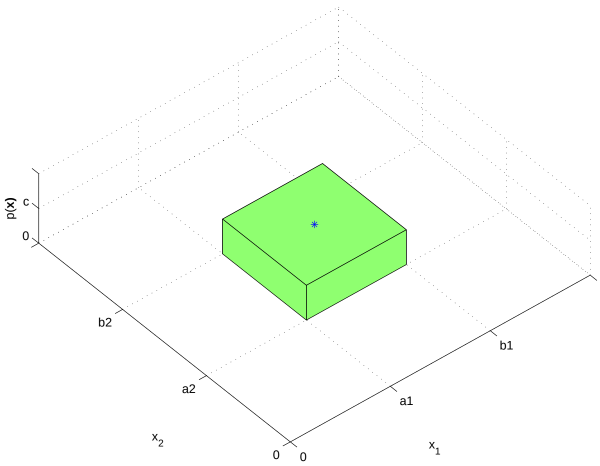

Figure 2: Uniform 2-dimensional probability density function with center of gravity shown raised to level c.

where

$$
\begin{aligned}
\mathbb{E}[x_1^2] &= \int_{a_2}^{b_2}\int_{a_1}^{b_1} x_1^2 \frac{1}{(b_1-a_1)(b_2-a_2)} dx_1 dx_2 \\
&= \frac{1}{(b_1-a_1)(b_2-a_2)} \frac{1}{3} [x_1^3]_{a_1}^{b_1} [x_2]_{a_2}^{b_2} \\
&= \frac{(b_2-a_2)(b_1^3-a_1^3)}{3(b_1-a_1)(b_2-a_2)} \\
&= \frac{(b_1-a_1)(b_1^2+a_1^2+a_1b_1)}{3(b_1-a_1)} \\
&= \frac{b_1^2+a_1^2+a_1b_1}{3}, \\
\mathbb{E}[x_2^2] &= \frac{b_2^2+a_2^2+a_2b_2}{3} \qquad (3)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{E}[x_1x_2] &= \frac{1}{(b_1-a_1)(b_2-a_2)} \int_{a_2}^{b_2} x_2 dx_2 \int_{a_1}^{b_1} x_1 dx_1 \\
&= \frac{1}{(b_1-a_1)(b_2-a_2)(2)(2)} [x_2^2]_{a_2}^{b_2} [x_1^2]_{a_1}^{b_1} \\
&= \frac{(b_1^2-a_1^2)(b_2^2-a_2^2)}{4(b_1-a_1)(b_2-a_2)} \\
&= \frac{(a_1+b_1)(a_2+b_2)}{4}. \qquad (4)
\end{aligned}
$$

Substituting this into (2) we get

$$
\mathbf{\Sigma} = \begin{bmatrix} \frac{(a_1-b_1)^2}{12} & 0 \\ 0 & \frac{(a_2-b_2)^2}{12}. \end{bmatrix} \qquad (5)
$$

d) All variables in $\boldsymbol{x}$ have the same variance.

e) The variables in $\boldsymbol{x}$ are uncorrelated.

Problem 5

a) We first find the eigenvalues by solving the characteristic equation given as

$$\begin{aligned} |(\boldsymbol{M} - \lambda \boldsymbol{I})| &= 0 \\ \begin{vmatrix} 5-\lambda & 3 \\ 3 & 5-\lambda \end{vmatrix} &= 0 \\ (5-\lambda)^2 - 9 &= 0 \\ &\Downarrow \\ \lambda_1 = 8 \qquad & \lambda_2 = 2. \end{aligned} \tag{1}$$

Then we find the corresponding eigenvectors by substituting the eigenvalues we found into $\boldsymbol{M}\boldsymbol{x} = \lambda\boldsymbol{x}$ and solve for $\boldsymbol{x}$.

Substitute $\lambda_1 = 8$:

$$\begin{aligned} \boldsymbol{M}\boldsymbol{x} &= \lambda_1 \boldsymbol{x} \\ (\boldsymbol{M} - \lambda_1 \boldsymbol{I})\boldsymbol{x} &= \boldsymbol{0} \\ \begin{pmatrix} 5-\lambda_1 & 3 \\ 3 & 5-\lambda_1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= 0 \\ \begin{pmatrix} -3 & 3 \\ 3 & -3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= 0 \\ -3x_1 \quad 3x_2 &= 0 \\ 3x_1 \quad -3x_2 &= 0 \\ &\Downarrow \\ x_2 = x_1 \qquad & x_1 \text{ arbitrary} \\ &\Downarrow \\ x_1 = t \qquad & x_2 = t \end{aligned} \tag{2}$$

.

Correspondingly we solve for $\lambda_2 = 2$:

$$\begin{aligned}
\boldsymbol{M}\boldsymbol{x} &= \lambda_2 \boldsymbol{x} \\
(\boldsymbol{M} - \lambda_2 \boldsymbol{I})\boldsymbol{x} &= \mathbf{0} \\
\begin{pmatrix} 5-\lambda_2 & 3 \\ 3 & 5-\lambda_2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= 0 \\
\begin{pmatrix} 3 & 3 \\ 3 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= 0 \\
3x_1 \quad 3x_2 &= 0 \\
3x_1 \quad 3x_2 &= 0 \\
&\Downarrow \\
x_2 = -x_1 \qquad & \qquad x_1 \text{ arbitrary} \\
&\Downarrow \\
x_1 = t \qquad & \qquad x_2 = -t
\end{aligned} \tag{3}$$

.

Thus we get the eigenvectors $\mathbf{e}_1 = \begin{pmatrix} t & t \end{pmatrix}^t$ and $\mathbf{e}_2 = \begin{pmatrix} t & -t \end{pmatrix}^t$ for λ_1 and λ_2 respectively.

We can substitute arbitrary variables, e.g. $\mathbf{e}_1 = 1/\sqrt{2}\begin{pmatrix} 1 & 1 \end{pmatrix}^t$ og $\mathbf{e}_2 = 1/\sqrt{2}\begin{pmatrix} -1 & 1 \end{pmatrix}^t$ (the scaling ensures unity length).

Thus we get the eigenvector- and eigenvalue-matrices $\mathbf{\Phi}$ and $\mathbf{\Lambda}$:

$$\begin{aligned}
\mathbf{\Phi} &= \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \\
\mathbf{\Lambda} &= \begin{pmatrix} 8 & 0 \\ 0 & 2 \end{pmatrix}
\end{aligned} \tag{4}$$

The principal axes are drawn as a vector between $\boldsymbol{\mu}$ and $\mathbf{p}_i = \boldsymbol{\mu} + \sqrt{\lambda_i}\mathbf{e}_i, i = 1, 2$ which are found to be $(3\ 3)^t$ and $(0\ 2)^t$ respectively.

b) The contour line of the probability density function are shown in figure 3 along with the principal axes.

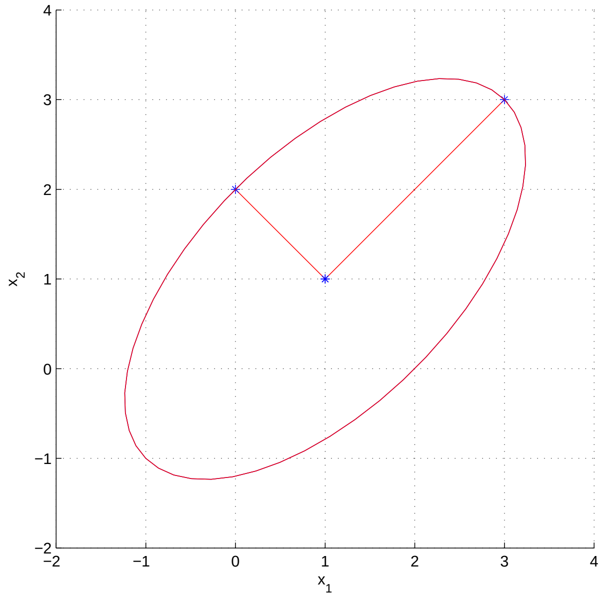

Figure 3: Contour line for 2D distribution.