Theoretical exercise 2

Håvard Godal

January 20, 2021

Problem 1

A company producing sea food plans to start using an automated classification system to detect possible toxic content in blue mussels. The classification will be based on chemical analyses of one mussel from each container in a batch of incoming containers of newly caught blue mussels, and the concentration x of a specific chemical substance in the mussels will be used as feature for classification. If this concentration is above a specific threshold value, the whole container will be considered poisonous and rejected. The company's gain for each container is NOK 250,-, but the company has also issued a guarantee to its customers that if they buy a container that shows itself to be poisonous the company will compensate costs up to NOK 100000,-.

a) - Let ω_1 = 'toxic mussels' og ω_2 = 'non toxic mussels'. Formulate the loss functions $\lambda(\alpha_i|\omega_j), i = 1, 2, j = 1, 2$, where α_i corresponds to the decision ω_i (risk will be the same as cost in this setting) when we do not include the possibility of rejecting classifications.

If non toxic mussles are sold and toxic mussles are not sold, the loss is 0. If a non toxic mussle is not sold, the loss will be 250kr. If a toxic mussle is sold, the loss will be 100000.

This results in the following loss functions:

$$\lambda(\alpha_1|\omega_1) = 0 \tag{1.1}$$

$$\lambda(\alpha_2|\omega_1) = 100000 \tag{1.2}$$

$$\lambda(\alpha_1|\omega_2) = 250 \tag{1.3}$$

$$\lambda(\alpha_2|\omega_2) = 0 \tag{1.4}$$

$$\boldsymbol{\lambda}(\alpha_i|\omega_j) = \begin{bmatrix} 0 & 250 \\ 100000 & 0 \end{bmatrix} \tag{1.5}$$

b) - Dependent on the value of x we classify as ω_1 or ω_2. Express the conditional loss functions, $R(\alpha_i|x)), i = 1, 2$.

We have that:

$$p(\boldsymbol{x}|\omega_1) \sim N(0.4, 0.0001) \tag{1.6}$$

$$p(\boldsymbol{x}|\omega_2) \sim N(0.2, 0.0001) \tag{1.7}$$

$$P(\omega_1) = \frac{1}{25} \tag{1.8}$$

$$P(\omega_2) = \frac{24}{25} \tag{1.9}$$

Furthermore, we can calculate the conditional loss functions like so:

$$R(\alpha_i|x) = \sum_{j=1}^{M} \lambda(\alpha_i|\omega_j) \cdot P(\omega_j|x) \tag{1.10}$$

Which results in:

$$\begin{aligned}
R(\alpha_1|x) &= \sum_{j=1}^{M} \lambda(\alpha_1|\omega_j) \cdot P(\omega_j|x) \\
&= \lambda(\alpha_1|\omega_1) \cdot P(\omega_1|x) + \lambda(\alpha_1|\omega_2) \cdot P(\omega_2|x) \\
&= 0 + 250 \cdot P(\omega_2|x) = 250 \cdot \frac{p(x|\omega_2) \cdot P(\omega_2)}{p(x)} \\
&= 250 \cdot P(\omega_2|x) = 250 \cdot \frac{p(x|\omega_2) \cdot P(\omega_2)}{P(\omega_1) \cdot p(x|\omega_1) + P(\omega_2) \cdot p(x|\omega_2)} \\
&= 250 \cdot \frac{\frac{1}{0.01 \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-0.2)^2}{0.0002}} \cdot \frac{24}{25}}{\frac{1}{25} \cdot \frac{1}{0.01 \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-0.4)^2}{0.0002}} + \frac{24}{25} \cdot \frac{1}{0.01 \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-0.2)^2}{0.0002}}} \\
&= \underline{\underline{\frac{6000 \cdot e^{-\frac{(x-0.2)^2}{0.0002}}}{e^{-\frac{(x-0.4)^2}{0.0002}} + 24 \cdot e^{-\frac{(x-0.2)^2}{0.0002}}}}}
\end{aligned} \tag{1.11}$$

and:

$$\begin{aligned}
R(\alpha_2|x) &= \sum_{j=1}^{M} \lambda(\alpha_2|\omega_j) \cdot P(\omega_j|x) \\
&= 100000 \cdot P(\omega_1|x) + 0 \\
&= 100000 \cdot \frac{p(x|\omega_1) \cdot P(\omega_1)}{p(x)} \\
&= \underline{\underline{\frac{100000 \cdot e^{-\frac{(x-0.4)^2}{0.0002}}}{e^{-\frac{(x-0.4)^2}{0.0002}} + 24 \cdot e^{-\frac{(x-0.2)^2}{0.0002}}}}}
\end{aligned} \tag{1.12}$$

c) - Determine the decision boundary that minimises the overall loss (average cost) upon classification (R).

The decision boundary occurs when the conditional loss functions are equal.

$$
\begin{aligned}
R(\alpha_1|x) &= R(\alpha_2|x) \\
6000 \cdot e^{-\frac{(x-0.2)^2}{0.0002}} &= 100000 \cdot e^{-\frac{(x-0.4)^2}{0.0002}} \\
\ln(6000) - \frac{(x-0.2)^2}{0.0002} &= \ln(100000) - \frac{(x-0.4)^2}{0.0002} \\
\frac{(x^2 - 0.8x + 0.16) - (x^2 - 0.4x + 0.04)}{0.0002} &= \ln(100000) - \ln(6000) \\
-2000x + 600 &= \ln(100000) - \ln(6000) \\
x &= \frac{\ln(100000) - \ln(6000) - 600}{-2000} \\
x &= x_0 = \underline{\underline{0.2986...}}
\end{aligned}
\tag{1.13}
$$

This means that for toxicity concentrations (x-values) smaller than x_0, the mussels are assumed to be non toxic, and toxic otherwise.

d) - Determine the minimum average cost R upon classification in NOK.

The minimum average cost can be found by integrating over the smallest probability density function given x.

$$
\begin{aligned}
R &= \int_{-\infty}^{x_0} R(\alpha_2|x) \cdot p(x) dx + \int_{x_0}^{\infty} R(\alpha_1|x) \cdot p(x) dx \\
&= \int_{-\infty}^{x_0} \lambda_{12} \cdot P(\omega_1) \cdot p(x|\omega_1) dx + \int_{x_0}^{\infty} \lambda_{21} \cdot P(\omega_2) \cdot p(x|\omega_2) dx \\
&= \frac{1}{25} 100000 \int_{-\infty}^{x_0} p(x|\omega_1) dx + \frac{24}{25} 250 \int_{x_0}^{\infty} p(x|\omega_2) dx \\
&= 4000 \cdot P(x < x_0|\omega_1) + 240 \cdot P(x > x_0|\omega_2) \\
&= 4000 \cdot P\left(z < \frac{0.2986 - 0.4}{0.01}\right) + 240 \cdot \left(1 - P\left(z < \frac{0.2986 - 0.2}{0.01}\right)\right) \\
&= 4000 \cdot P(z < -10.14) + 240 \cdot (1 - P(z < 9.86)) \\
&\approx 4000 \cdot 0 + 240 \cdot (1 - 1) \\
&\approx \underline{\underline{0}}
\end{aligned}
\tag{1.14}
$$

This means that the probability density functions have almost no overlap, and the chance of misclassification is approximately zero.

Problem 2

We want to design a pattern recognition system that classifies 2-dimensional feature vectors $\{\boldsymbol{x}\}$ to class ω_1, or class ω_2.

It is known that the distribution between the two classes are 1/2 and 1/2 respectively for class ω_1 and class ω_2. Furthermore the feature vectors of the two classes are normally distributed around $\mu_1 = (3\ 3)^T$ with covariance matrix

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 1/2 & 0 \\ 0 & 2 \end{bmatrix} \tag{2.1}$$

for class ω_1, og around $\mu_2 = (3\ -2)^T$ with covariance matrix

$$\boldsymbol{\Sigma}_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \tag{2.2}$$

for class ω_2

The eigenvalue and eigenvector matrices $\boldsymbol{\Lambda}_i$ and $\boldsymbol{\Phi}_i$, $i = 1, 2$ for the convariance matrices are

$$\boldsymbol{\Lambda}_1 = \begin{bmatrix} 1/2 & 0 \\ 0 & 2 \end{bmatrix} \tag{2.3}$$

and

$$\boldsymbol{\Phi}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{2.4}$$

for class ω_1, and

$$\boldsymbol{\Lambda}_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \tag{2.5}$$

and

$$\boldsymbol{\Phi}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{2.6}$$

for class ω_2 respectively.

a) - Sketch the contour lines for the class specific probability density functions for the two classes in the same diagram.

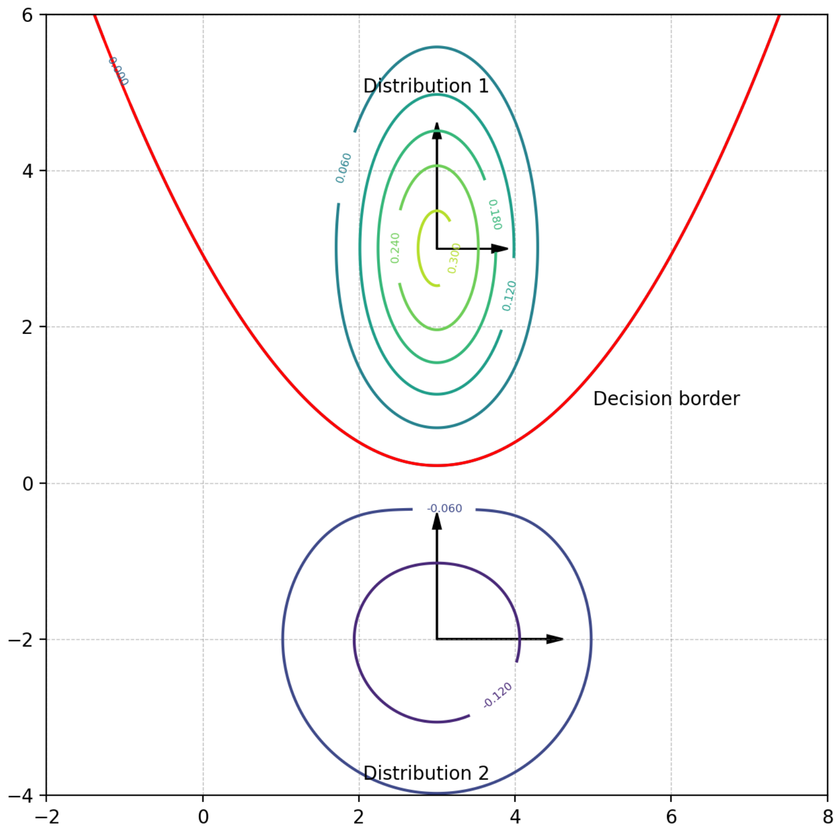

Figure 1: Contour lines for the class conditional density functions

b) - Formulate Bayes decision rule for this problem. Determine the decision boundary and decision regions in the same diagram as in a).

Bayes decision rule is as follows:

$$\text{Decide} \begin{cases} \omega_1 & \text{if } P(\omega_1|\boldsymbol{x}) > P(\omega_2|\boldsymbol{x}) \\ \omega_2 & \text{otherwise} \end{cases} \tag{2.7}$$

And from the exercise notes:

$$g_i(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{\Theta}_i\boldsymbol{x} + \boldsymbol{\theta}_i^T\boldsymbol{x} + \theta_{i0} \tag{2.8}$$

where:

$$\begin{aligned} \boldsymbol{\Theta}_i &= -\frac{1}{2}\Sigma_i^{-1} \\ \boldsymbol{\theta}_i &= \Sigma_i^{-1}\boldsymbol{\mu}_i \\ \theta_{i0} &= -\frac{1}{2}\mu_i^T\Sigma_i^{-1}\mu_i - \frac{1}{2}\ln|\Sigma_i| + \ln(P(\omega_i)) \end{aligned} \tag{2.9}$$

Solving for $g_1(\boldsymbol{x})$:

$$\begin{aligned} \boldsymbol{\Theta}_1 &= -\frac{1}{2}\begin{bmatrix} 2 & 0 \\ 0 & 1/2 \end{bmatrix} \\ &= \begin{bmatrix} -1 & 0 \\ 0 & -1/4 \end{bmatrix} \end{aligned} \tag{2.10}$$

$$
\begin{aligned}
\boldsymbol{\theta}_1 &= \begin{bmatrix} 2 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} 3 \\ 3 \end{bmatrix} \\
&= \begin{bmatrix} 6 \\ 3/2 \end{bmatrix}
\end{aligned} \tag{2.11}
$$

$$
\begin{aligned}
\theta_{10} &= -\frac{1}{2} \begin{bmatrix} 3 & 3 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} 3 \\ 3 \end{bmatrix} - \frac{1}{2} \ln(1/2 \cdot 2 + 0 \cdot 0) + \ln(1/2) \\
&= -11.943
\end{aligned} \tag{2.12}
$$

$$
\begin{aligned}
g_1(\boldsymbol{x}) &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & -1/4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 6 & 3/2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - 11.943 \\
&= -x_1^2 - \frac{1}{4} x_2^2 + 6x_1 + \frac{3}{2} x_2 - 11.943
\end{aligned} \tag{2.13}
$$

Solving for $g_2(\boldsymbol{x})$:

$$
\begin{aligned}
\boldsymbol{\Theta}_2 &= -\frac{1}{2} \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix} \\
&= \begin{bmatrix} -1/4 & 0 \\ 0 & -1/4 \end{bmatrix}
\end{aligned} \tag{2.14}
$$

$$
\begin{aligned}
\boldsymbol{\theta}_2 &= \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} 3 \\ 2- \end{bmatrix} \\
&= \begin{bmatrix} 3/2 \\ -1 \end{bmatrix}
\end{aligned} \tag{2.15}
$$

$$
\begin{aligned}
\theta_{20} &= -\frac{1}{2} \begin{bmatrix} 3 & -2 \end{bmatrix} \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \end{bmatrix} - \frac{1}{2} \ln(2 \cdot 2 + 0 \cdot 0) + \ln(1/2) \\
&= -4.636
\end{aligned} \tag{2.16}
$$

$$
\begin{aligned}
g_2(\boldsymbol{x}) &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} -1/4 & 0 \\ 0 & -1/4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} -2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - 4.636 \\
&= -\frac{1}{4} x_1^2 - \frac{1}{4} x_2^2 + \frac{3}{2} x_1 - x_2 - 4.636
\end{aligned} \tag{2.17}
$$

The decision border occurs when $g_1(\boldsymbol{x}) = g_2(\boldsymbol{x})$

$$
\begin{aligned}
g_1(\boldsymbol{x}) &= g_2(\boldsymbol{x}) \\
-x_1^2 + 6x_1 - \frac{1}{4} x_2^2 + \frac{3}{2} x_2 - 11.943 &= -\frac{1}{4} x_1^2 - \frac{1}{4} x_2^2 + \frac{3}{2} x_1 - x_2 - 4.636 \\
-\frac{3}{4} x_1^2 + \frac{9}{2} x_1 + \frac{5}{2} x_2 - 7.307 &= 0 \\
x_2 &= \frac{3}{10} x_1^2 - \frac{9}{5} x_1 + 2.923
\end{aligned} \tag{2.18}
$$

Problem 3

Show that with

$$g_i(\boldsymbol{x}) = -\frac{\|\boldsymbol{x} - \boldsymbol{\mu}_i\|^2}{2\sigma^2} + \ln P(\omega_i) \tag{3.1}$$

as a starting point where $i = 1, 2$, we can reach an expression defining the decision border between the two classes:

$$\boldsymbol{\theta}^T(\boldsymbol{x} - \boldsymbol{x}_0) = 0 \tag{3.2}$$

where

$$\boldsymbol{\theta} = \boldsymbol{\mu}_i - \boldsymbol{\mu}_j \tag{3.3}$$

and

$$\boldsymbol{x}_0 = \frac{1}{2}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j) - \frac{\sigma^2}{\|\boldsymbol{\mu}_i + \boldsymbol{\mu}_j\|^2} \ln \frac{P(\omega_i)}{P(\omega_j)}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j) \tag{3.4}$$

From the lecture notes, the following calculations can be done with $g_i(\boldsymbol{x})$:

$$\begin{aligned} g_i(\boldsymbol{x}) &= -\frac{1}{2\sigma^2} (\boldsymbol{x} - \boldsymbol{\mu}_i)^T (\boldsymbol{x} - \boldsymbol{\mu}_i) + \ln(P(\omega_i)) \\ &= -\frac{1}{2\sigma^2} \left[\boldsymbol{x}^T\boldsymbol{x} - 2\boldsymbol{\mu}_i^T\boldsymbol{x} + \boldsymbol{\mu}_i^T\boldsymbol{\mu}_i\right] + \ln(P(\omega_i)) \\ &\text{Removing the class independent terms} \\ &= -\frac{1}{2\sigma^2} \left[-2\boldsymbol{\mu}_i^T\boldsymbol{x} + \boldsymbol{\mu}_i^T\boldsymbol{\mu}_i\right] + \ln(P(\omega_i)) \end{aligned} \tag{3.5}$$

Setting $g_i(\boldsymbol{x}) = g_j(\boldsymbol{x})$

$$\begin{aligned} -\frac{1}{2\sigma^2} \left[-2\boldsymbol{\mu}_i^T\boldsymbol{x} + \boldsymbol{\mu}_i^T\boldsymbol{\mu}_i\right] + \ln(P(\omega_i)) &= -\frac{1}{2\sigma^2} \left[-2\boldsymbol{\mu}_j^T\boldsymbol{x} + \boldsymbol{\mu}_j^T\boldsymbol{\mu}_j\right] + \ln(P(\omega_j)) \\ -\frac{1}{2\sigma^2} \left[-2\boldsymbol{\mu}_i^T\boldsymbol{x} + \boldsymbol{\mu}_i^T\boldsymbol{\mu}_i\right] + \frac{1}{2\sigma^2} \left[-2\boldsymbol{\mu}_j^T\boldsymbol{x} + \boldsymbol{\mu}_j^T\boldsymbol{\mu}_j\right] + \ln(P(\omega_i)) - \ln(P(\omega_j)) &= 0 \\ -\frac{1}{2} \left[-2\boldsymbol{\mu}_i^T\boldsymbol{x} + \boldsymbol{\mu}_i^T\boldsymbol{\mu}_i\right] + \frac{1}{2} \left[-2\boldsymbol{\mu}_j^T\boldsymbol{x} + \boldsymbol{\mu}_j^T\boldsymbol{\mu}_j\right] + \sigma^2 \ln\left(\frac{P(\omega_i)}{P(\omega_j)}\right) &= 0 \\ \left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right)^T \boldsymbol{x} - \frac{1}{2}\boldsymbol{\mu}_i^T\boldsymbol{\mu}_i + \frac{1}{2}\boldsymbol{\mu}_j^T\boldsymbol{\mu}_j + \sigma^2 \ln\left(\frac{P(\omega_i)}{P(\omega_j)}\right) &= 0 \\ \text{Expanding the expression with } \frac{1}{2}\left(\boldsymbol{\mu}_i^T\boldsymbol{\mu}_j - \boldsymbol{\mu}_j^T\boldsymbol{\mu}_i\right) \text{ (which equals 0)} & \end{aligned}$$

$$\left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right)^T \boldsymbol{x} - \frac{1}{2}\left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right)^T\left(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j\right) + \sigma^2 \ln\left(\frac{P(\omega_i)}{P(\omega_j)}\right) = 0 \tag{3.6}$$

$$\begin{aligned} &\text{To write the expression in the desired form, the last term is extended.} \\ &\left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right)^T \boldsymbol{x} - \frac{1}{2}\left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right)^T\left(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j\right) + \sigma^2 \ln\left(\frac{P(\omega_i)}{P(\omega_j)}\right) \frac{\left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right)^T\left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right)}{\left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right)^T\left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right)} = 0 \\ &\text{Rewriting the denominator and moving equal terms outside the parentheses.} \\ &\left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right)^T \left[\boldsymbol{x} - \left(\frac{1}{2}\left(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j\right) - \frac{\sigma^2}{\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|^2} \ln\frac{P(\omega_i)}{P(\omega_j)}\left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right)\right)\right] = 0 \\ &\text{Which equals to:} \\ &\boldsymbol{\theta}^T(\boldsymbol{x} - \boldsymbol{x}_0) = 0 \end{aligned}$$