Machine Learning - Theoretical exercise 2

Téo Bouvard

January 20, 2020

Problem 1

a) In the following, we use the notation $\lambda(\alpha_i \mid \omega_j) \Leftrightarrow \lambda_{ij}$

$\lambda_{11} = 0$	correct classification of toxic container
$\lambda_{21} = 10^5$	incorrect classification of toxic container
$\lambda_{22} = 0$	correct classification of non-toxic container
$\lambda_{12} = 250$	incorrect classification of non-toxic container

b)

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

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

$$R(\alpha_1 \mid x) = \lambda_{12} P(\omega_2 \mid x)$$
$$R(\alpha_2 \mid x) = \lambda_{21} P(\omega_1 \mid x)$$

c) To determine the decision boundary that minimizes the average cost, we solve the equality of conditional loss functions.

$$R(\alpha_1 \mid x) = R(\alpha_2 \mid x)$$
$$\lambda_{12} P(\omega_2 \mid x) = \lambda_{21} P(\omega_1 \mid x)$$
$$\frac{\lambda_{12}}{\lambda_{21}} \frac{P(\omega_2) P(x \mid \omega_2)}{P(x)} = \frac{P(\omega_1) P(x \mid \omega_1)}{P(x)} \qquad \text{(Bayes Theorem)}$$
$$\frac{\lambda_{12} P(\omega_2)}{\lambda_{21} P(\omega_1)} P(x \mid \omega_2) = P(x \mid \omega_1)$$

Let $K = \frac{\lambda_{12} P(\omega_2)}{\lambda_{21} P(\omega_1)}$. Furthermore, we know that $P(x \mid \omega_1) \sim \mathcal{N}(\mu_1, \sigma^2)$ and $P(x \mid \omega_2) \sim \mathcal{N}(\mu_2, \sigma^2)$.

$$\begin{aligned}
K \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu_1)^2}{2\sigma^2}} &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu_2)^2}{2\sigma^2}} \\
\ln{(K e^{-\frac{(x-\mu_1)^2}{2\sigma^2}})} &= \ln{(e^{-\frac{(x-\mu_2)^2}{2\sigma^2}})} \\
\ln K - \frac{(x-\mu_1)^2}{2\sigma^2} &= -\frac{(x-\mu_2)^2}{2\sigma^2} \\
\ln K + \frac{1}{2\sigma^2}((x-\mu_2)^2 - (x-\mu_1)^2) &= 0 \\
\ln K + \frac{1}{2\sigma^2}(x^2 - 2\mu_2 x + \mu_2^2 - x^2 + 2\mu_1 x - \mu_1^2) &= 0 \\
\ln K + \frac{1}{2\sigma^2}(2x(\mu_1 - \mu_2) + \mu_2^2 - \mu_1^2) &= 0 \\
\ln K + \frac{\mu_1 - \mu_2}{\sigma^2} x + \frac{\mu_2^2 - \mu_1^2}{2\sigma^2} &= 0 \\
\frac{\mu_1 - \mu_2}{\sigma^2} x &= \frac{\mu_1^2 - \mu_2^2}{2\sigma^2} - \ln K \\
x &= \frac{\mu_1^2 - \mu_2^2}{2(\mu_1 - \mu_2)} - \sigma^2 \ln K \\
x &= \frac{\mu_1 + \mu_2}{2} - \frac{\sigma^2}{\mu_1 - \mu_2} \ln K
\end{aligned}$$

Numerically solving this equation gives us the decision boundary

$$\begin{aligned}
x_0 &= \frac{0.4 + 0.2}{2} - \frac{10^{-4}}{0.4 - 0.2} \times \ln(\frac{25 \times 250}{10^5}) \\
x_0 &= 0.3014
\end{aligned}$$

d) We determine the minimum average cost R upon classification

$$\begin{aligned}
R &= \int_{-\infty}^{x_0} R(\alpha_2 \mid x) P(x) dx + \int_{x_0}^{+\infty} R(\alpha_1 \mid x) P(x) dx \\
R &= \int_{-\infty}^{x_0} \lambda_{21} P(\omega_1 \mid x) P(x) dx + \int_{x_0}^{+\infty} \lambda_{12} P(\omega_2 \mid x) P(x) dx \\
&= \lambda_{21} \int_{-\infty}^{x_0} \frac{P(\omega_1) P(x \mid \omega_1)}{P(x)} P(x) dx + \lambda_{12} \int_{x_0}^{+\infty} \frac{P(\omega_2) P(x \mid \omega_2)}{P(x)} P(x) dx \\
&= \lambda_{21} P(\omega_1) \int_{-\infty}^{x_0} P(x \mid \omega_1) dx + \lambda_{12} P(\omega_2) \int_{x_0}^{+\infty} P(x \mid \omega_2) dx \\
&= \lambda_{21} P(\omega_1) P(x < x_0 \mid \omega_1) + \lambda_{12} P(\omega_2) P(x > x_0 \mid \omega_2)
\end{aligned}$$

To get $P(\omega_1)$ and $P(\omega_2)$, we use the fact that

$$\begin{cases} P(\omega_1) + P(\omega_2) & = 1 \\ \frac{P(\omega_2)}{P(\omega_1)} & = 25 \end{cases} \implies \begin{cases} P(\omega_1) & = \frac{1}{26} \\ P(\omega_2) & = \frac{25}{26} \end{cases}$$

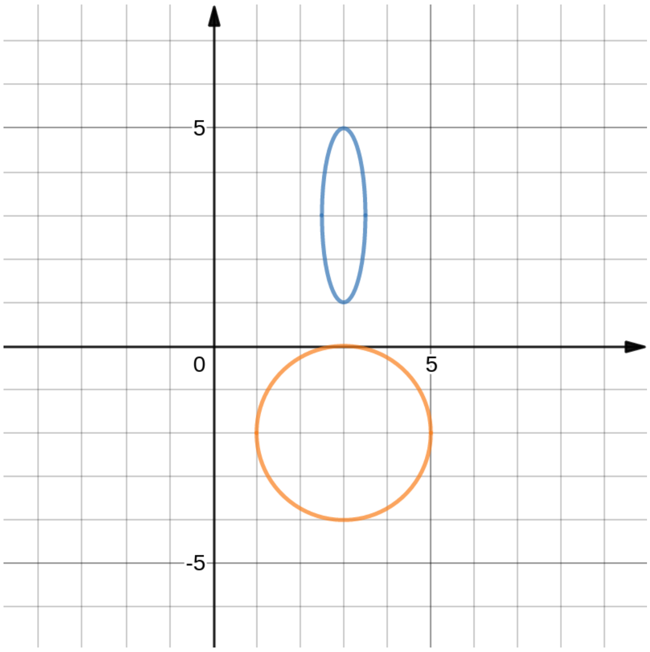
Using a standard normal table,

$$\begin{aligned}
P(x < x_0 \mid \omega_1) = P(y < \frac{x_0 - \mu_1}{\sigma} \mid \omega_1) \approx 10^{-23} \\
P(x > x_0 \mid \omega_2) = P(y < \frac{x_0 - \mu_2}{\sigma} \mid \omega_2) \approx 10^{-23}
\end{aligned}$$

Which gives us a minimum average cost of $R = 10^{-20}$ NOK. Because the variance σ is so small for both distributions, the average estimated risk is nearly 0.

Problem 2

a)



b) The risk function g can be defined as $g_i(x) = \ln P(\omega_i) + \ln P(x \mid \omega_i)$ for $i = 1, 2$. In this exercise, priors are equal so they can be omitted when solving the equality of risk for both classes.

$$
\begin{aligned}
g_1(x) &= g_2(x) \\
\ln P(\omega_1) + \ln P(x \mid \omega_1) &= \ln P(\omega_2) + \ln P(x \mid \omega_2) \\
\ln P(x \mid \omega_1) &= \ln P(x \mid \omega_2)
\end{aligned}
$$

Furthermore, the feature vectors of the two classes are normally distributed, meaning that

$$
P(x \mid \omega_i) = \frac{1}{(2\pi)^{\frac{d}{2}} \left|\Sigma_i\right|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)} \qquad\qquad i = 1, 2
$$

So the previous equality becomes

$$
\begin{aligned}
-\frac{d}{2}\ln 2\pi - \frac{1}{2}\ln\left|\Sigma_1\right| - \frac{1}{2}(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1) &= -\frac{d}{2}\ln 2\pi - \frac{1}{2}\ln\left|\Sigma_2\right| - \frac{1}{2}(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2) \\
-\frac{1}{2}\ln\left|\Sigma_1\right| - \frac{1}{2}(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1) &= -\frac{1}{2}\ln\left|\Sigma_2\right| - \frac{1}{2}(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)
\end{aligned}
$$

The covariance matrices Σ_i have nice properties allowing us to simplify this equation.

$$
\left|\Sigma_1\right| = \begin{vmatrix} \frac{1}{2} & 0 \\ 0 & 2 \end{vmatrix} = 1 \implies \frac{1}{2}\ln\left|\Sigma_1\right| = 0 \tag{1}
$$

$$
\left|\Sigma_2\right| = \begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix} = 4 \implies \frac{1}{2}\ln\left|\Sigma_2\right| = \ln 2 \tag{2}
$$

$$
\Sigma_1^{-1} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \qquad \text{because } \Sigma_1 \text{ is diagonal} \tag{3}
$$

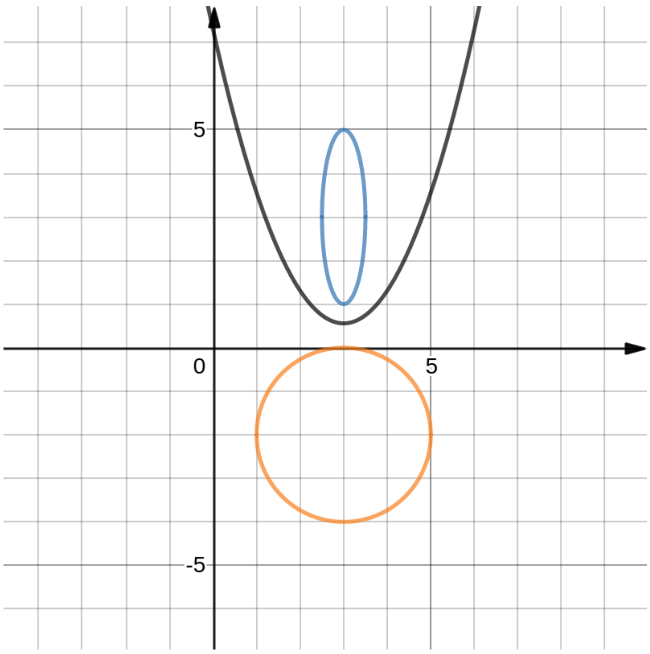
$$
\Sigma_2^{-1} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \qquad \text{because } \Sigma_2 \text{ is diagonal} \tag{4}
$$

This allows us to write the previous equality as

$$
\begin{aligned}
-\frac{1}{2}(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1)+\frac{1}{2}(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)+\ln 2 &= 0 \\
-\frac{1}{4}\begin{bmatrix} x_1-3 \\ x_2-3 \end{bmatrix}^T \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1-3 \\ x_2-3 \end{bmatrix} + \frac{1}{4}\begin{bmatrix} x_1-3 \\ x_2+2 \end{bmatrix}^T \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1-3 \\ x_2+2 \end{bmatrix} + \ln 2 &= 0 \\
-\frac{1}{4}[(4x_1-12)(x_1-3)+(x_2-3)^2]+\frac{1}{4}[(x_1-3)^2+(x_2+2)^2]+\ln 2 &= 0 \\
-\frac{1}{4}(4x_1^2-24x_1+36+x_2^2-6x_2+9)+\frac{1}{4}(x_1^2-6x_1+9+x_2^2+4x_2+4)+\ln 2 &= 0 \\
\frac{3}{4}x_1^2+\frac{9}{2}x_1-8+\ln 2 &= x_2
\end{aligned}
$$

Thus, the decision border is a parabola.

c)



Problem 3

$$
\begin{aligned}
g_i(x) &= g_j(x) \\
-\frac{\|x-\mu_i\|^2}{2\sigma^2}+\ln P(\omega_1) &= -\frac{\|x-\mu_j\|^2}{2\sigma^2}+\ln P(\omega_2) \\
\frac{\|x-\mu_j\|^2-\|x-\mu_i\|^2}{2\sigma^2}+\ln\frac{P(\omega_1)}{P(\omega_2)} &= 0
\end{aligned}
$$

By remarking that $\|u\|^2 = u^T u$, we get

$$
\begin{aligned}
\frac{(\mu_i - \mu_j)^T(2x - (\mu_i + \mu_j))}{2\sigma^2} + \ln\frac{P(\omega_1)}{P(\omega_2)} &= 0 \\
\frac{2(\mu_i - \mu_j)^T x - (\mu_i - \mu_j)^T(\mu_i + \mu_j)}{2\sigma^2} + \ln\frac{P(\omega_1)}{P(\omega_2)} &= 0 \\
\frac{(\mu_i - \mu_j)^T x}{\sigma^2} - \frac{(\mu_i - \mu_j)^T(\mu_i + \mu_j)}{2\sigma^2} + \ln\frac{P(\omega_1)}{P(\omega_2)} &= 0 \\
(\mu_i - \mu_j)^T x - \frac{1}{2}(\mu_i - \mu_j)^T(\mu_i + \mu_j) + \sigma^2 \ln\frac{P(\omega_1)}{P(\omega_2)} &= 0 \\
(\mu_i - \mu_j)^T(x - \frac{1}{2}(\mu_i + \mu_j) + \frac{\sigma^2}{(\mu_i - \mu_j)^T}\ln\frac{P(\omega_1)}{P(\omega_2)}) &= 0 \\
(\mu_i - \mu_j)^T(x - \frac{1}{2}(\mu_i + \mu_j) + \frac{\sigma^2}{\left\|\mu_i - \mu_j\right\|^2}\ln\frac{P(\omega_1)}{P(\omega_2)}(\mu_i - \mu_j)) &= 0
\end{aligned}
$$

By assigning

$$\theta = \mu_i - \mu_j \tag{5}$$

$$x_0 = \frac{1}{2}(\mu_i + \mu_j) - \frac{\sigma^2}{\left\|\mu_i - \mu_j\right\|^2}\ln\frac{P(\omega_1)}{P(\omega_2)}(\mu_i - \mu_j)) \tag{6}$$

we get the following linear equation

$$\theta^T(x - x_0) = 0$$