Machine Learning 1: Assignment 1

Tom Nick 340528
Niklas Gebauer xxxxx
Felix Bohmann xxxxxx

Exercise 1

a)

$$P(error) = \int P(error \mid x)p(x)dx \tag{1}$$
$$P(error \mid x) = \min(P(w_1 \mid x), P(w_2 \mid x)) \tag{2}$$

With these equations, we want to show that

$$P(error) \leq \int \frac{2}{\frac{1}{P(w_1|x)} + \frac{1}{P(w_2|x)}} p(x)dx \tag{3}$$

At first, without restricting the general case, we assume that $P(w_1 \mid x) \geq P(w_2 \mid x)$, that is the function $P(error \mid x) = P(w_2 \mid x)$. Now with (**??**), (**??**) and (**??**) we have:

$$\int P(w_2 \mid x)p(x)dx \leq \int \frac{2}{\frac{1}{P(w_1|x)} + \frac{1}{P(w_2|x)}} p(x)dx$$

Because both sides are integrating over the same variable we can simplify the term to:

$$\begin{aligned}
&P(w_2 \mid x)p(x) \leq \frac{2}{\frac{1}{P(w_1|x)} + \frac{1}{P(w_2|x)}} p(x) \\
&\Leftrightarrow P(w_2 \mid x) \leq \frac{2}{\frac{1}{P(w_1|x)} + \frac{1}{P(w_2|x)}} \\
&\Leftrightarrow (\frac{1}{P(w_1 \mid x)} + \frac{1}{P(w_2 \mid x)})P(w_2 \mid x) \leq 2 \\
&\Leftrightarrow \frac{1}{P(w_1 \mid x)} + \frac{1}{P(w_2 \mid x)} \leq \frac{2}{P(w_2 \mid x)} \\
&\Leftrightarrow \frac{1}{P(w_1 \mid x)} \leq \frac{1}{P(w_2 \mid x)} \\
&\Leftrightarrow P(w_1 \mid x) \geq P(w_2 \mid x)
\end{aligned}$$

This holds true with the assumptions we made earlier.

b)

With this result, we now show that:

$$P(error) \leq \frac{2P(w_1)P(w_2)}{\sqrt{P(w_1)^2 + (4\mu^2 + 2)P(w_1)P(w_2) + P(w_2)^2}}$$

While using the univariate probability distribution:

$$p(x \mid w_1) = \frac{\pi^{-1}}{1 + (x - \mu)^2} \text{ and } p(x \mid w_2) = \frac{\pi^{-1}}{1 + (x + \mu)^2}$$

With the rule of bayes we have $P(w_1 \mid x) = \frac{p(x|w_1)P(w_1)}{p(x)}$:

$$
\begin{aligned}
&P(error) \leq \int \frac{2}{\frac{1}{P(w_1|x)} + \frac{1}{P(w_2|x)}} p(x)dx \\
\Leftrightarrow\ &P(error) \leq \int \frac{2}{\frac{1}{\frac{p(x|w_1)P(w_1)}{p(x)}} + \frac{1}{\frac{p(x|w_2)P(w_2)}{p(x)}}} p(x)dx \\
\Leftrightarrow\ &P(error) \leq \int \frac{2}{\frac{p(x)}{p(x|w_1)P(w_1)} + \frac{p(x)}{p(x|w_2)P(w_2)}} p(x)dx \\
\Leftrightarrow\ &P(error) \leq \int \frac{2}{\frac{1}{p(x|w_1)P(w_1)} + \frac{1}{p(x|w_2)P(w_2)}} dx \\
\Leftrightarrow\ &P(error) \leq \int \frac{2}{\frac{1}{\frac{\pi^{-1}}{1+(x-\mu)^2}P(w_1)} + \frac{1}{\frac{\pi^{-1}}{1+(x+\mu)^2}P(w_2)}} dx \\
\Leftrightarrow\ &P(error) \leq \int \frac{2}{\frac{1}{\frac{\pi^{-1}P(w_1)}{1+(x-\mu)^2}} + \frac{1}{\frac{\pi^{-1}P(w_2)}{1+(x+\mu)^2}}} dx \\
\Leftrightarrow\ &P(error) \leq \int \frac{2}{\frac{1+(x-\mu)^2}{\pi^{-1}P(w_1)} + \frac{1+(x+\mu)^2}{\pi^{-1}P(w_2)}} dx \\
\Leftrightarrow\ &P(error) \leq \int \frac{2}{\frac{(1+(x-\mu)^2)P(w_2)}{\pi^{-1}P(w_1)P(w_2)} + \frac{(1+(x+\mu)^2)P(w_1)}{\pi^{-1}P(w_2)P(w_1)}} dx \\
\Leftrightarrow\ &P(error) \leq \int \frac{2\pi^{-1}P(w_2)P(w_1)}{(1+(x+\mu)^2)P(w_1) + (1+(x-\mu)^2)P(w_2)} dx \\
\Leftrightarrow\ &P(error) \leq \int \frac{2\pi^{-1}P(w_2)P(w_1)}{(x^2+2x\mu+\mu^2+1)P(w_1) + (x^2-2x\mu+\mu^2+1)P(w_2)} dx \\
\Leftrightarrow\ &P(error) \leq \int \frac{2\pi^{-1}P(w_2)P(w_1)}{(P(w_1)+P(w_2))x^2 + (P(w_1)-P(w_2))2\mu x + (P(w_1)\mu^2 + P(w_2)\mu^2 + P(w1) + P(w_2))} dx
\end{aligned}
$$

We can now take out the numerator of the integral and use the following equation:

$$\int \frac{1}{ax^2+bx+c} dx = \frac{2\pi}{\sqrt{4ac-b^2}} \tag{4}$$

with:

$$
\begin{aligned}
a &= P(w_1) + P(w_2) \\
b &= (P(w_1) - P(w_2))2\mu \\
c &= P(w_1)\mu^2 + P(w_2)\mu^2 + P(w1) + P(w_2)
\end{aligned}
$$

because

$$
\begin{aligned}
& b^2 < 4ac \\
\Leftrightarrow\ & 0 < 4ac - b^2 \\
\Leftrightarrow\ & 0 < 4(P(w_1) + P(w_2))(P(w_1)\mu^2 + P(w_2)\mu^2 + P(w1) + P(w_2)) - ((P(w_1) - P(w_2))2\mu)^2 \\
\Leftrightarrow\ & 0 < 4P(w_1)^2\mu^2 + 4P(w_1)P(w_2)\mu^2 + 4P(w_1)^2 + 4P(w_1)P(w_2) \\
& \quad + 4P(w_1)P(w_2)\mu^2 + 4P(w_2)^2\mu^2 + 4P(w_2)P(w_1) + 4P(w_2)^2 \\
& \quad - (4P(w_1)^2\mu^2 - 8P(w_1)P(w_2)\mu^2 + 4P(w_2)^2\mu^2) \\
\Leftrightarrow\ & 0 < 16P(w_1)P(w_2)\mu^2 + 8P(w_1)P(w_2) + 4P(w_1)^2 + 4P(w_2)^2
\end{aligned}
$$

and this holds since $P(w_1), P(w_2) \in [0, 1]$ and $P(w_1) + P(w_2) = 1$.
We now already calculated $4ac - b^2$ in the steps before and just need to use (4) to proceed where we stopped before introducing equation 4:

$$
\begin{aligned}
\Leftrightarrow P(error) &\leq \int \frac{2\pi^{-1}P(w_2)P(w_1)}{(P(w_1) + P(w_2))x^2 + (P(w_1) - P(w_2))2\mu x + (P(w_1)\mu^2 + P(w_2)\mu^2 + P(w1) + P(w_2))} dx \\
\Leftrightarrow P(error) &\leq 2\pi^{-1}P(w_2)P(w_1)\frac{2\pi}{\sqrt{16P(w_1)P(w_2)\mu^2 + 8P(w_1)P(w_2) + 4P(w_1)^2 + 4P(w_2)^2}} \\
\Leftrightarrow P(error) &\leq \frac{4P(w_1)P(w_2)}{\sqrt{4((4\mu^2 + 2)P(w_1)P(w_2) + P(w_1)^2 + P(w_2)^2)}} \\
\Leftrightarrow P(error) &\leq \frac{4P(w_1)P(w_2)}{\sqrt{4}\sqrt{((4\mu^2 + 2)P(w_1)P(w_2) + P(w_1)^2 + P(w_2)^2)}} \\
\Leftrightarrow P(error) &\leq \frac{2P(w_1)P(w_2)}{\sqrt{P(w_1)^2 + (4\mu^2 + 2)P(w_1)P(w_2) + P(w_2)^2}}
\end{aligned}
$$

c)

According to informations form chapter 2.8 in Pattern Classification:
For both case we can find the upper-bound numericaly via the Chernoff Bound or the Bhattacharyya Bound, this may lead to a tighter upper bound than the analytical version, which is even harder to approximate due to the discontinues nature of the integrals. For the high-dimensional space the Bhattacharyya Bound might be better, because it is computationally less expensive that the Chernoff Bound.

Exercise 2

a)

The data is generated by the univariate Laplacian Distrubution:

$$
p(x \mid w_1) = \frac{1}{2\sigma}\exp\left(-\frac{\mid x - \mu \mid}{\sigma}\right) \text{ and } p(x \mid w_2) = \frac{1}{2\sigma}\exp\left(-\frac{\mid x + \mu \mid}{\sigma}\right)
$$

To get the optimal decision boundary we have to solve $P(w_1 \mid x) = P(w_2 \mid x)$ for x.

$$
\begin{aligned}
& & P(w_1 \mid x) &= P(w_2 \mid x) \\
&\Leftrightarrow & \frac{p(x \mid w_1)P(w_1)}{p(x)} &= \frac{p(x \mid w_2)P(w_2)}{p(x)} \\
&\Leftrightarrow & p(x \mid w_1)P(w_1) &= p(x \mid w_2)P(w_2) \\
&\Leftrightarrow & \frac{1}{2\sigma}\exp\left(-\frac{\mid x-\mu \mid}{\sigma}\right)P(w_1) &= \frac{1}{2\sigma}\exp\left(-\frac{\mid x+\mu \mid}{\sigma}\right)P(w_2) \\
&\Leftrightarrow & \exp\left(-\frac{\mid x-\mu \mid}{\sigma}\right)P(w_1) &= \exp\left(-\frac{\mid x+\mu \mid}{\sigma}\right)P(w_2) \\
&\Leftrightarrow & \ln(\exp\left(-\frac{\mid x-\mu \mid}{\sigma}\right)P(w_1)) &= \ln(\exp\left(-\frac{\mid x+\mu \mid}{\sigma}\right)P(w_2)) \\
&\Leftrightarrow & -\frac{\mid x-\mu \mid}{\sigma} + \ln(P(w_1)) &= -\frac{\mid x+\mu \mid}{\sigma} + \ln(P(w_2)) \\
&\Leftrightarrow & \ln(P(w_1)) - \ln(P(w_2)) &= \frac{\mid x-\mu \mid}{\sigma} - \frac{\mid x+\mu \mid}{\sigma} \\
&\Leftrightarrow & \ln\left(\frac{P(w_1)}{P(w_2)}\right) &= \frac{\mid x-\mu \mid - \mid x+\mu \mid}{\sigma} \\
&\Leftrightarrow & \sigma\ln\left(\frac{P(w_1)}{P(w_2)}\right) &= \mid x-\mu \mid - \mid x+\mu \mid
\end{aligned}
$$

There are now 3 cases to look at:

1. $x \leq \mu$ and $-x \leq \mu$

$$
\begin{aligned}
\sigma\ln\left(\frac{P(w_1)}{P(w_2)}\right) &= -(x-\mu) - (x+\mu) \\
&= -2x \\
-\frac{\sigma}{2}\ln\left(\frac{P(w_1)}{P(w_2)}\right) &= x
\end{aligned}
$$

2. $x \leq \mu$ and $-x > \mu$

$$
\begin{aligned}
\sigma\ln\left(\frac{P(w_1)}{P(w_2)}\right) &= -(x-\mu) + (x+\mu) \\
&= 2\mu
\end{aligned}
$$

3. $x > \mu$

$$
\begin{aligned}
\sigma\ln\left(\frac{P(w_1)}{P(w_2)}\right) &= (x-\mu) - (x+\mu) \\
&= -2\mu
\end{aligned}
$$

b)

To satisfy the constrain of $P(w_2 \mid x) = P(error \mid x)$, $P(w_1 \mid x)$ must contain $P(w_2 \mid x)$. Visually it would mean that the area of the curve of $P(w_2 \mid x)$ is completly included inside the area of the curve of $P(w_1 \mid x)$. When varying the mean μ, both curves move away from each other, which means it always has to stay 0, else it's not possible to satisfy the constrain. The scale σ can take any value that is larger than zero, it has the same effect on both distributions. $P(w_1)$ must always be larger than $P(w_2)$ (or else $P(w_1 \mid x)$ can't contain $P(w_2 \mid x)$).

Formally all possible value-tuples are:

$$\{(P(w_1), P(w_2), \mu, \sigma) \mid P(w_1) \geq P(w_2) \wedge \mu = 0 \wedge \sigma \in \mathbb{R}^+\}$$

c)

The data is generated by the univariate Gaussian Distrubution:

$$p(x \mid w_1) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \text{ and } p(x \mid w_2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x+\mu)^2}{2\sigma^2}\right)$$

To get the optimal decision boundary we have to solve $P(w_1 \mid x) = P(w_2 \mid x)$ for x.

$$\begin{aligned}
& & P(w_1 \mid x) &= P(w_2 \mid x) \\
\Leftrightarrow & & \frac{p(x \mid w_1)P(w_1)}{p(x)} &= \frac{p(x \mid w_2)P(w_2)}{p(x)} \\
\Leftrightarrow & & \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) P(w_1) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x+\mu)^2}{2\sigma^2}\right) P(w_2) \\
\Leftrightarrow & & \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) P(w_1) &= \exp\left(-\frac{(x+\mu)^2}{2\sigma^2}\right) P(w_2) \\
\Leftrightarrow & & 2\sigma^2 \ln\left(\frac{P(w_1)}{P(w_2)}\right) &= x^2 - 2x\mu + \mu^2 - x^2 - 2x\mu - \mu^2 \\
\Leftrightarrow & & 2\sigma^2 \ln\left(\frac{P(w_1)}{P(w_2)}\right) &= -4x\mu \\
\Leftrightarrow & & \frac{2\sigma^2}{-4\mu} \ln\left(\frac{P(w_1)}{P(w_2)}\right) &= x
\end{aligned}$$