

Machine Learning 1

Exercise Sheet 1

TU Berlin
WS 2017/18

Group: BSSBCH

Matthias Bigalke, 339547, maku@win.tu-berlin.de

Tolga Buz, 346836, buz_tolga@yahoo.de

Alejandro Hernandez, 395678, alejandrohernandezmunuera@gmail.com

Aitor Palacios Cuesta, 396276, aitor.palacioscuesta@campus.tu-berlin.de

Christof Schubert, 344450, christof.schubert@campus.tu-berlin.de

Daniel Steinhaus, 342563, dany.steinhaus@googlemail.com

Exercise 1: Estimating Bayes Error

(a) Show that the full error can be upper-bounded as follows:

$$\begin{aligned} P(\text{error}) &\leq \int \frac{2}{\frac{1}{P(\omega_1|x)} + \frac{1}{P(\omega_2|x)}} p(x) dx \\ &\Leftrightarrow \int P(\text{error}|x)p(x)dx \leq \int \frac{2}{\frac{1}{P(\omega_1|x)} + \frac{1}{P(\omega_2|x)}} p(x) dx \\ &\Leftrightarrow \int P(\text{error}|x) p(x) dx \leq \int \frac{2 \cdot P(\omega_1|x) \cdot P(\omega_2|x)}{P(\omega_2|x) + P(\omega_1|x)} p(x) dx \end{aligned}$$

Using Bayes formula $P(\omega_i|x) = \frac{p(x|\omega_i) \cdot P(\omega_i)}{p(x)}$ leads to:

$$\begin{aligned} &\Leftrightarrow \int P(\text{error}|x)p(x)dx \leq \int \frac{2 \cdot P(\omega_1|x) \cdot P(\omega_2|x) \cdot p(x)}{\frac{p(x|\omega_1) \cdot P(\omega_1)}{p(x)} + \frac{p(x|\omega_2) \cdot P(\omega_2)}{p(x)}} dx \\ &\Leftrightarrow \int P(\text{error}|x)p(x)dx \leq \int \frac{2 \cdot P(\omega_1|x) \cdot P(\omega_2|x) \cdot p(x)^2}{p(x|\omega_1) \cdot P(\omega_1) + p(x|\omega_2) \cdot P(\omega_2)} dx \end{aligned}$$

With $p(x|\omega_1) \cdot P(\omega_1) + p(x|\omega_2) \cdot P(\omega_2) = p(x)$:

$$\begin{aligned} &\Leftrightarrow \int P(\text{error}|x)p(x)dx \leq \int 2 \cdot P(\omega_1|x) \cdot P(\omega_2|x) \cdot p(x) dx \\ &\Leftrightarrow \int \min[P(\omega_1|x), P(\omega_2|x)]p(x)dx \leq \int 2 \cdot P(\omega_1|x) \cdot P(\omega_2|x) \cdot p(x) dx \end{aligned}$$

Since $f(x) \leq g(x) \forall x$, then $\int f(x) dx \leq \int g(x) dx \forall x$ applies, we show for arbitrary x :

$$\min[P(\omega_1|x), P(\omega_2|x)]p(x) \leq 2 \cdot P(\omega_1|x) \cdot P(\omega_2|x) \cdot p(x)$$

Now there are two cases to consider which can be handled the same way. First we are looking at case $P(\omega_1|x) \leq P(\omega_2|x)$:

$$P(\omega_1|x) \cdot p(x) \leq 2 \cdot P(\omega_1|x) \cdot P(\omega_2|x)p(x)$$

$$\Leftrightarrow 1 \leq 2 \cdot P(\omega_2|x)$$

This is true, since $P(\omega_2|x) \geq P(\omega_1|x)$ and $P(\omega_1|x) + P(\omega_2|x) = 1$. Thus $P(\omega_2|x) \geq 0.5$ and $\Leftrightarrow 1 \leq 2 \cdot P(\omega_2|x)$ is valid.

The second case for $P(\omega_1|x) \geq P(\omega_2|x)$ can be solved analogously (and $1 \leq 2 \cdot P(\omega_1|x)$ is valid if $P(\omega_1|x) \geq P(\omega_2|x)$).

This shows that the full error can be upper-bound as given.

(b) Show that $P(\text{error}) \leq \frac{2P(\omega_1)P(\omega_2)}{\sqrt{1+4\mu^2 P(\omega_1)P(\omega_2)}}$

We are starting with the term of 1 a) which has been shown to be true:

$$\begin{aligned}
P(\text{error}) &\leq \int \frac{2}{\frac{1}{P(\omega_1|x)} + \frac{1}{P(\omega_2|x)}} p(x) dx \\
&\Leftrightarrow P(\text{error}) \leq \int \frac{2p(x)}{\frac{p(x)}{p(x|\omega_1)P(\omega_1)} + \frac{p(x)}{p(x|\omega_2)P(\omega_2)}} dx \\
&\Leftrightarrow P(\text{error}) \leq \int \frac{2}{\frac{1}{p(x|\omega_1)P(\omega_1)} + \frac{1}{p(x|\omega_2)P(\omega_2)}} dx \\
&\Leftrightarrow P(\text{error}) \leq \int \frac{2 \cdot p(x|\omega_1)P(\omega_1) \cdot p(x|\omega_2)P(\omega_2)}{p(x|\omega_1)P(\omega_1) + p(x|\omega_2)P(\omega_2)} dx \\
&\Leftrightarrow P(\text{error}) \leq 2 \cdot P(\omega_1) \cdot P(\omega_2) \int \frac{1}{\pi^2 \cdot (1+(x-\mu)^2)(1+(x+\mu)^2) \cdot (\frac{P(\omega_1)}{\pi(1+(x-\mu)^2)} + \frac{P(\omega_2)}{\pi(1+(x+\mu)^2)})} dx \\
&\Leftrightarrow P(\text{error}) \leq 2 \cdot P(\omega_1) \cdot P(\omega_2) \int \frac{1}{\pi(P(\omega_1) \cdot (1+(x+\mu)^2) + P(\omega_2) \cdot (1+(x-\mu)^2))} dx \\
&\Leftrightarrow P(\text{error}) \leq \frac{2 \cdot P(\omega_1) \cdot P(\omega_2)}{\pi} \int \frac{1}{P(\omega_1) + P(\omega_1)(x^2+2\mu x+\mu^2) + P(\omega_2) + P(\omega_2)(x^2-2\mu x+\mu^2)} dx \\
&\Leftrightarrow P(\text{error}) \leq \frac{2 \cdot P(\omega_1)P(\omega_2)}{\pi} \cdot \int \frac{1}{(P(\omega_1)+P(\omega_2))x^2 + (2\mu(P(\omega_1)-P(\omega_2))x + \mu^2 P(\omega_1) + \mu^2 P(\omega_2) + P(\omega_1) + P(\omega_2))} dx
\end{aligned}$$

Using the identity $\int \frac{1}{ax^2+bx+c} dx = \frac{2\pi}{\sqrt{4ac-b^2}}$ *:

$$\begin{aligned}
&\Leftrightarrow P(\text{error}) \leq \frac{2 \cdot P(\omega_1)P(\omega_2)}{\pi} \cdot \frac{2\pi}{\sqrt{4 \cdot (P(\omega_1)+P(\omega_2)) \cdot (1+\mu^2)(P(\omega_1)+P(\omega_2)) - 4\mu^2(P(\omega_1)-P(\omega_2))^2}} \\
&\Leftrightarrow P(\text{error}) \leq \frac{2 \cdot P(\omega_1)P(\omega_2)}{\pi} \cdot \frac{2\pi}{2\sqrt{(P(\omega_1)+P(\omega_2)) \cdot (1+\mu^2)(P(\omega_1)+P(\omega_2)) - 4\mu^2(P(\omega_1)-P(\omega_2))^2}} \\
&\Leftrightarrow P(\text{error}) \leq \frac{2 \cdot P(\omega_1)P(\omega_2)}{\sqrt{(P(\omega_1)+P(\omega_2))^2 + 2 \cdot P(\omega_1)P(\omega_2) + P(\omega_2)^2 + 4\mu^2 P(\omega_1)P(\omega_2)}} \\
&\Leftrightarrow P(\text{error}) \leq \frac{2 \cdot P(\omega_1)P(\omega_2)}{\sqrt{(P(\omega_1)+P(\omega_2))^2 + 4\mu^2 P(\omega_1)P(\omega_2)}}
\end{aligned}$$

With $(P(\omega_1) + P(\omega_2))^2 = 1$:

$$\Leftrightarrow P(\text{error}) \leq \frac{2 \cdot P(\omega_1)P(\omega_2)}{\sqrt{1+4\mu^2 P(\omega_1)P(\omega_2)}} \quad q.e.d.$$

(c) In low dimensional case we would use numerical integration for error estimation by sampling from the posterior probability over the input space.

For higher dimensional cases numerical integrations are mostly impracticable. We would estimate the error by using the Chernoff bound, because Chernoff bound is always tighter to the real error than Bhattacharyya bound. Another benefit of using this bound is the dimensionality reduction from arbitrarily high dimensional space to one dimension β space.

Exercise 2: Bayes Decision Boundaries

(a) The decision boundary will be:

$$\begin{aligned}
P(\omega_1|x) &= P(\omega_2|x) \\
\Leftrightarrow \frac{p(x|\omega_1) \cdot P(\omega_1)}{p(x)} &= \frac{p(x|\omega_2) \cdot P(\omega_2)}{p(x)} \\
\Leftrightarrow \frac{y}{2\sigma} \cdot \exp\left(\frac{-|x-\mu|}{\sigma}\right) \cdot P(\omega_1) &= \frac{y}{2\sigma} \cdot \exp\left(\frac{-|x+\mu|}{\sigma}\right) \cdot P(\omega_2)
\end{aligned}$$

*For using the integration identity formula, $b^2 \leq 4ac$ has to be ensured. This is given in our case, since $(1 + 4\mu^2 P(\omega_1)P(\omega_2)) > 0$

$$\Leftrightarrow -\frac{|x-\mu|}{\sigma} + \ln(P(\omega_1)) = -\frac{|x+\mu|}{\sigma} + \ln(P(\omega_2))$$

$$\Leftrightarrow \ln(P(\omega_1)) - \ln(P(\omega_2)) = \frac{|x-\mu|}{\sigma} - \frac{|x+\mu|}{\sigma}$$

To decrease the amounts the following four cases have to be considered and are stating:

- First Case: $(x + \mu) \geq 0$ and $(x - \mu) \geq 0$ or $(x \geq \mu)$:
 $(\ln(P(\omega_1)) - \ln(P(\omega_2))) \cdot \sigma = (x - \mu) - (x + \mu) = -2\mu$
 $\Leftrightarrow -(\ln(P(\omega_1)) - \ln(P(\omega_2))) \cdot \frac{\sigma}{2} = \mu$
- Second Case: $(x + \mu) \geq 0$ and $(x - \mu) \leq 0$ or $(-\mu \leq x \leq \mu)$:
 $(\ln(P(\omega_1)) - \ln(P(\omega_2))) \cdot \sigma = -(x - \mu) - (x + \mu) = -2x$
 $\Leftrightarrow -(\ln(P(\omega_1)) - \ln(P(\omega_2))) \cdot \frac{\sigma}{2} = x$
- Third Case: $(x + \mu) \leq 0$ and $(x - \mu) \leq 0$ or $(x \leq -\mu)$:
 $(\ln(P(\omega_1)) - \ln(P(\omega_2))) \cdot \sigma = -(x - \mu) + (x + \mu) = 2\mu$
 $\Leftrightarrow (\ln(P(\omega_1)) - \ln(P(\omega_2))) \cdot \frac{\sigma}{2} = \mu$
- The fourth case $(x + \mu) \leq 0$ and $(x - \mu) \geq 0$ is not possible since $\mu > 0$ by definition.

So it follows, that the optimal decision boundary is $-(\ln(P(\omega_1)) - \ln(P(\omega_2))) \cdot \frac{\sigma}{2} = x$

(b) Determine for which $P(\omega_1)$, $P(\omega_2)$, μ and σ the optional decision is to always predict the first class.

For this the $P(\omega_1)$, $P(\omega_2)$, μ and σ have to be found for which holds $P(\omega_1|x) > P(\omega_2|x) \forall x \in \mathbb{R}$.

If $P(\omega_1) = 1$, $\mu > 0$ and $\sigma > 0$ are arbitrary.

In all other cases, starting from the initial inequality, we get:

$$\begin{aligned} P(\omega_1|x) &\geq P(\omega_2|x) \\ \Leftrightarrow \frac{p(x|\omega_1) \cdot P(\omega_1)}{p(x)} &\geq \frac{p(x|\omega_2) \cdot P(\omega_2)}{p(x)} \\ \Leftrightarrow \frac{y}{2\sigma} \cdot \exp\left(\frac{-|x-\mu|}{\sigma}\right) \cdot P(\omega_1) &\geq \frac{y}{2\sigma} \cdot \exp\left(\frac{-|x+\mu|}{\sigma}\right) \cdot P(\omega_2) \\ \Leftrightarrow -\frac{|x-\mu|}{\sigma} + \ln(P(\omega_1)) &\geq -\frac{|x+\mu|}{\sigma} + \ln(P(\omega_2)) \\ \Leftrightarrow \ln(P(\omega_1)) - \ln(P(\omega_2)) &\geq \frac{|x-\mu|}{\sigma} - \frac{|x+\mu|}{\sigma} \end{aligned}$$

To decrease the amounts the following four cases have to be considered and are stating the decision boundaries:

- First Case: $(x + \mu) \geq 0$ and $(x - \mu) \geq 0$ or $(x \geq \mu)$:
 $\Leftrightarrow (\ln(P(\omega_1)) - \ln(P(\omega_2))) \cdot \sigma \geq (x - \mu) - (x + \mu) = -2\mu$
 $\Leftrightarrow -(\ln(P(\omega_1)) - \ln(P(\omega_2))) \cdot \left(\frac{\sigma}{2}\right) \leq \mu$
- Second Case: $(x + \mu) \geq 0$ and $(x - \mu) \leq 0$ or $(-\mu \leq x \leq \mu)$:
 $\Leftrightarrow (\ln(P(\omega_1)) - \ln(P(\omega_2))) \cdot \sigma \geq -(x - \mu) - (x + \mu) = -2x$
 $\Leftrightarrow -(\ln(P(\omega_1)) - \ln(P(\omega_2))) \cdot \frac{\sigma}{2} \leq x$
- Third Case: $(x + \mu) \leq 0$ and $(x - \mu) \leq 0$ or $(x \leq -\mu)$:
 $\Leftrightarrow (\ln(P(\omega_1)) - \ln(P(\omega_2))) \cdot \sigma \geq -(x - \mu) - (-(x + \mu)) = 2\mu$
 $\Leftrightarrow (\ln(P(\omega_1)) - \ln(P(\omega_2))) \cdot \frac{\sigma}{2} \geq \mu$
- The fourth case $(x + \mu) \leq 0$ and $(x - \mu) \geq 0$ is not possible since $\mu > 0$ by definition.

Considering now the inequalities for $(x \geq \mu)$, $(-\mu \leq x \leq \mu)$ and $(x \leq -\mu)$ which have to be met for predicting the first class over the second:

If $-(\ln(P(\omega_1)) - \ln(P(\omega_2))) \cdot \left(\frac{\sigma}{2}\right) \leq \mu \leq (\ln(P(\omega_1)) - \ln(P(\omega_2))) \cdot \frac{\sigma}{2}$ all inequalities are fulfilled for their definition ranges for every x .

Since $\mu > 0$ by definition, the following condition has to be met for always predicting the first class:
 $\mu \leq (\ln(P(\omega_1)) - \ln(P(\omega_2))) \cdot \frac{\sigma}{2}$

(c) Boundary for Gaussian distribution:

$$P(\omega_1|x) = P(\omega_2|x)$$

$$\Leftrightarrow \ln(P(\omega_1)) - \frac{(x-\mu)^2}{2\sigma^2} = \ln(P(\omega_2)) - \frac{(x+\mu)^2}{2\sigma^2}$$

$$\Leftrightarrow \ln(P(\omega_1)) - \ln(P(\omega_2)) = \frac{(x-\mu)^2 - (x+\mu)^2}{2\sigma^2}$$

$$\Leftrightarrow (\ln(P(\omega_1)) - \ln(P(\omega_2))) \cdot 2\sigma^2 = x^2 - 2x\mu + \mu^2 - (x^2 + 2x\mu + \mu^2)$$

$$\Leftrightarrow (\ln(P(\omega_1)) - \ln(P(\omega_2))) \cdot 2\sigma^2 = -4x\mu$$

$$\Leftrightarrow x = -(\ln(P(\omega_1)) - \ln(P(\omega_2))) \cdot \frac{\sigma^2}{2\mu}$$

$$\text{So the boundary is given by } x = -(\ln(P(\omega_1)) - \ln(P(\omega_2))) \cdot \frac{\sigma^2}{2\mu}.$$

Values so that $\forall x : P(w_1|x) \geq P(w_2|x)$

If $P(\omega_1) = 1$, μ and $\sigma > 0$ are arbitrary. In all other cases μ and σ can't be found.