# Machine Learning 1

## Exercise Sheet 1

(09. November 2020)

**Students:**    Mattes Ohse                    337356
Florian Ebert                   391837
Rodrigo Alexis Pardo Meza
Bertty Contreras Rojas

# 1 Exercise 1: Estimating the Bayes Error

## 1.1 a

The Bayes Formular is given by:

$$P(w_j) = \frac{P(x \mid w_j)}{p(x)}. \tag{1}$$

and the Bayes Error is given by:

$$P(error) = min[P(w_1|x), P(w_2|x)]. \tag{2}$$

We now wanna show, that the full error can be upper-bounded as:

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

We can assume that $P(w_1|x) \geq P(w_2|x)$, so that (2) becomes to $P(error) = P(w_2|x)$. Now we can use (3), under consideration that we are integrating on both sides for the same variable what means that we can ignore the integralsigns:

$$\Rightarrow \qquad P(error) \leq \frac{2}{\frac{1}{P(w_1|x)} + \frac{1}{P(w_2|x)}} p(x)$$

$$\Leftrightarrow \qquad P(w_2|x)p(x) \leq \frac{2}{\frac{1}{P(w_1|x)} + \frac{1}{P(w_2|x)}} p(x)$$

$$\Leftrightarrow \qquad P(w_2|x) \leq \frac{2}{\frac{1}{P(w_1|x)} + \frac{1}{P(w_2|x)}}$$

$$\Leftrightarrow \qquad \frac{P(w_2|x)}{P(w_1|x)} + \frac{P(w_2|x)}{P(w_2|x)} \leq 2$$

$$\Leftrightarrow \qquad \frac{1}{P(w_1|x)} + \frac{1}{P(w_2|x)} \leq \frac{2}{P(w_2|x)}$$

$$\Leftrightarrow \qquad \frac{1}{P(w_1|x)} \leq \frac{1}{P(w_2|x)}$$

$$\Leftrightarrow \qquad P(w_1|x) \geq P(w_2|x).$$

And that improves (3)

## 1.2 b

We want to show that for the distributions

$$p(x|w_1) = \frac{\pi^{-1}}{1 + (x - \mu)^2}, \ p(x|w_2) = \frac{\pi^{-1}}{1 + (x + \mu)^2} \tag{4}$$

that the Bayes error can be upper-bounded by:

$$P(error) \leq \frac{2P(w_1)P(w_2)}{\sqrt{1 + 4\mu^2 P(w_1)P(w_2)}} \tag{5}$$

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

$$= \int \frac{2}{\frac{p(x)}{p(x|w_1) \cdot P(w_1)} + \frac{p(x)}{p(x|w_2) \cdot P(w_2)}} p(x) dx$$

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

$$= \int \frac{2}{\frac{1}{\frac{\pi^{-1}}{1+(x-\mu)^2} P(w_1)} + \frac{1}{\frac{\pi^{-1}}{1+(x+\mu)^2} P(w_2)}} dx$$

$$= \int \frac{2}{\frac{\pi((x-\mu)^2+1)}{P(w_1)} + \frac{\pi((x+\mu)^2+1)}{P(w_2)}} dx$$

$$= \int \frac{2}{\frac{\pi(P(w_2)((x-\mu)^2+1)+P(w_1)((x+\mu)^2+1))}{P(w_1)P(w_2)}}$$

$$= \int \frac{2 \cdot P(w_1)P(w_2)}{\pi(P(w_2)((x-\mu)^2+1) + P(w_1)((x+\mu)^2+1))}$$

$$= \frac{2 \cdot P(w_1)P(w_2)}{\pi} \int \frac{1}{P(w_2)(x-\mu)^2 + P(w_2) + P(w_1)(x+\mu)^2 + P(w_1)} dx$$

Use $(x - \mu)^2 = x^2 - 2x\mu + \mu^2$ and $(x + \mu)^2 = x^2 + 2x\mu + \mu^2$

2

$$=\frac{2 \cdot P(w_1)P(w_2)}{\pi}$$

$$\cdot \int \frac{dx}{x^2 P(w_2) - 2x\mu P(w_2) + \mu^2 P(w_2) + P(w_2) + x^2 P(w_1) + 2x\mu P(w_1) + \mu^2 P(w_1) + P(w_1)}$$

$$=\frac{2 \cdot P(w_1)P(w_2)}{\pi}$$

$$\cdot \int \frac{dx}{x^2(P(w_2) + P(w_1)) + x2\mu(P(w_1) - P(w_2)) + \mu^2(P(w_2) + P(w_1)) + P(w_2) + P(w_1)}$$

We say:
$a = P(w_1) + P(w_2),$
$b = 2\mu(P(w_1 - P(w_2))$
$c = \mu^2(P(w_1) + P(w_2)) + P(w_1) + P(w_2)$
and we use: $\displaystyle\int \frac{1}{ax^2 + bx + c}dx = \frac{2\pi}{\sqrt{4ac - b^2}}.$

Now we take a look at 4ac:
$$4ac = 4[P(w_1) + P(w_2)] \cdot [\mu^2(P(w_1) + P(w_2)) + P(w_1) + P(w_2)]$$
$$= 4(\mu^2 + 1)[P(w_1) + P(w_2)]^2.$$
With that and the result of $b^2$
$$b^2 = 4\mu(P(w_1) - P(w_2))^2 = 4\mu^2[P^2(w_1) - 2P(w_1)P(w_2) + P^2(w_2)]$$
becomes the integral above to:

$$\Rightarrow \frac{2P(w_1)P(w_2)}{\pi}$$

$$\cdot \frac{2\pi}{\sqrt{4(\mu^2+1)[P(w_1)+P(w_2)]^2 - 4\mu^2[P^2(w_1) - 2P(w_1)P(w_2) + P^2(w_2)]}}$$

$$= \frac{2P(w_1)P(w_2)}{\sqrt{4\mu^2 P(w_1)P(w_2) + 2P(w_1)P(w_2) + P(w_1) + P(w_2)}} \Big| (\text{use: } (a+b)^2 = a^2 + 2ab + b^2)$$

$$= \frac{2P(w_1)P(w_2)}{\sqrt{4\mu^2 P(w_1)P(w_2) + (P(w_1) + P(w_2))^2}} \Big| (\text{We know that: } P(w_1) + P(w_2) = 1)$$

$$= \frac{2P(w_1)P(w_2)}{\sqrt{1 + 4\mu^2 P(w_1)P(w_2)}}$$

This proves that the full error can be upper-bounded as:

$$P(error) \leq \frac{2P(w_1)P(w_2)}{\sqrt{1 + 4\mu^2 P(w_1)P(w_2)}} \tag{6}$$

## 1.3 c

For the two cases whether the data is low-dimensional or the data is high-dimensional we can find the upper-bounds by using the methods Chernoff Bound or Bhattacharyya Bound. Both of these methods are numerical. The functions of these integrals are usually not continuous, so that the bounds determined in this way are often tighter than the bounds determined analytically. In high-dimensional cases, Bhattacharyya's method should be better because it is less computationally expensive than the Chernoff Bound method.

# 2 Exercise 2: Bayes Decision Boundaries

## 2.1 a

Univariate Laplacian Distrubution:

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

The optimal decision is to always predict the first class if $\forall x \in \mathbb{R} : P(w_1 \mid x) > P(w_2 \mid x)$.

$$P(w_1 \mid x) > P(w_2 \mid x)$$

$$\Leftrightarrow \quad \frac{p(x \mid w_1)P(w_1)}{p(x)} > \frac{p(x \mid w_2)P(w_2)}{p(x)}$$

$$\Leftrightarrow \quad p(x \mid w_1)P(w_1) > p(x \mid w_2)P(w_2)$$

$$\Leftrightarrow \quad \frac{1}{2\sigma} \exp\left(-\frac{\mid x - \mu \mid}{\sigma}\right) P(w_1) > \frac{1}{2\sigma} \exp\left(-\frac{\mid x + \mu \mid}{\sigma}\right) P(w_2)$$

$$\Leftrightarrow \quad \exp\left(-\frac{\mid x - \mu \mid}{\sigma}\right) P(w_1) > \exp\left(-\frac{\mid x + \mu \mid}{\sigma}\right) P(w_2)$$

$$\Leftrightarrow \quad \ln\left(\exp\left(-\frac{\mid x - \mu \mid}{\sigma}\right) P(w_1)\right) > \ln\left(\exp\left(-\frac{\mid x + \mu \mid}{\sigma}\right) P(w_2)\right)$$

$$\Leftrightarrow \quad -\frac{\mid x - \mu \mid}{\sigma} + \ln(P(w_1)) > -\frac{\mid x + \mu \mid}{\sigma} + \ln(P(w_2))$$

$$\Leftrightarrow \quad \ln(P(w_1)) - \ln(P(w_2)) > \frac{\mid x - \mu \mid}{\sigma} - \frac{\mid x + \mu \mid}{\sigma}$$

$$\Leftrightarrow \quad \ln\left(\frac{P(w_1)}{P(w_2)}\right) > \frac{\mid x - \mu \mid - \mid x + \mu \mid}{\sigma}$$

$$\Leftrightarrow \quad \sigma \ln\left(\frac{P(w_1)}{P(w_2)}\right) > \mid x - \mu \mid - \mid x + \mu \mid$$

Three cases for the absolute values:

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

$$\sigma \ln\left(\frac{P(w_1)}{P(w_2)}\right) > -(x - \mu) - (x + \mu)$$

$$> -2x$$

$$-\frac{\sigma}{2} \ln\left(\frac{P(w_1)}{P(w_2)}\right) < x$$

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

$$\sigma \ln \left( \frac{P(w_1)}{P(w_2)} \right) > -(x - \mu) + x + \mu$$
$$> 2\mu$$

3. $x > \mu$

$$\sigma \ln \left( \frac{P(w_1)}{P(w_2)} \right) > (x - \mu) - (x + \mu)$$
$$> -2\mu$$

Possible value-tuples where always the first class will be choosen:

$$\{(P(w_1), P(w_2), \mu, \sigma) \,|(P(w_1) > P(w_2))$$
$$\wedge (P(w_1) = 1 - P(w_2))$$
$$\wedge \sigma > 0 \wedge \mu > 0$$
$$\wedge ((\sigma \ln \left( \frac{P(w_1)}{P(w_2)} \right) > -2x \wedge x \leq \mu \wedge -x \leq \mu)$$
$$\vee (\sigma \ln \left( \frac{P(w_1)}{P(w_2)} \right) > 2\mu \wedge x \leq \mu \wedge -x > \mu)$$
$$\vee (\sigma \ln \left( \frac{P(w_1)}{P(w_2)} \right) > -2\mu) \wedge x > \mu)\}$$

## 2.2 b

Univariate Gaussian Distrubution:

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

The optimal decision is to always predict the first class if $\forall x \in \mathbb{R}: P(w_1 \mid x) > P(w_2 \mid x)$.

$$P(w_1 \mid x) > P(w_2 \mid x)$$

$$\Leftrightarrow \quad \frac{p(x \mid w_1)P(w_1)}{p(x)} > \frac{p(x \mid w_2)P(w_2)}{p(x)}$$

$$\Leftrightarrow \quad \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) P(w_1) > \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x+\mu)^2}{2\sigma^2}\right) P(w_2)$$

$$\Leftrightarrow \quad \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) P(w_1) > \exp\left(-\frac{(x+\mu)^2}{2\sigma^2}\right) P(w_2)$$

$$\Leftrightarrow \quad -\frac{(x-\mu)^2}{2\sigma^2} + \ln\left(P(w_1)\right) > -\frac{(x+\mu)^2}{2\sigma^2} + \ln\left(P(w_2)\right)$$

$$\Leftrightarrow \quad \ln\left(\frac{P(w_1)}{P(w_2)}\right) > \frac{(x-\mu)^2 - (x+\mu)^2}{2\sigma^2}$$

$$\Leftrightarrow \quad 2\sigma^2 \ln\left(\frac{P(w_1)}{P(w_2)}\right) > x^2 - 2x\mu + \mu^2 - x^2 - 2x\mu - \mu^2$$

$$\Leftrightarrow \quad 2\sigma^2 \ln\left(\frac{P(w_1)}{P(w_2)}\right) > -4x\mu$$

$$\Leftrightarrow \quad \frac{2\sigma^2}{-4\mu} \ln\left(\frac{P(w_1)}{P(w_2)}\right) < x$$

Possible value-tuples where always the first class will be choosen:

$$\{(P(w_1), P(w_2), \mu, \sigma) \mid (P(w_1) > P(w_2))$$
$$\wedge (P(w_1) = 1 - P(w_2))$$
$$\wedge\, \sigma > 0 \wedge \mu > 0$$
$$\wedge\, (\frac{2\sigma^2}{-4\mu} \ln\left(\frac{P(w_1)}{P(w_2)}\right) < x)\}$$