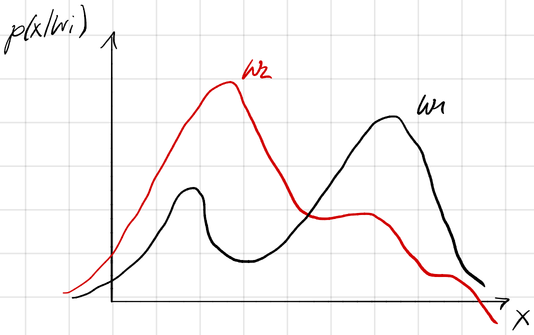
Sheet 1 Theory

Exercise 1: Estimating the Bayes Error

$$P(error) = \int P(error|x)\, p(x)\, dx$$

, where $\quad P(error|x) = \min[P(w_1|x), P(w_2|x)]$



$$P(error|x) = \begin{cases} P(w_1|x) & \text{if we decide } w_2 \\ P(w_2|x) & \text{if we decide } w_1 \end{cases}$$

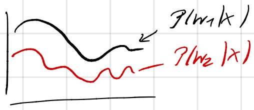
a) Show that the full error $P(error)$ can be upper-bounded as follows

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

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

$$P(error|x) = \min[P(w_1|x), P(w_2|x)]$$

Assume : $P(w_1|x) \ge P(w_2|x)$



$\Rightarrow \quad P(error|x) = \min[\ldots] = P(w_2|x)$

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

integration over same variable

$$P(w_2|x)\, p(x) \le \frac{2}{\frac{1}{P(w_1|x)} + \frac{1}{P(w_2|x)}}\, p(x)$$

$P(w_1) + P(w_2) = 1$

$$\left[\frac{1}{P(w_1|x)} + \frac{1}{P(w_2|x)}\right] P(w_2|x) \le 2$$

$$\frac{1}{P(w_1|x)} + \frac{1}{P(w_2|x)} \le \frac{2}{P(w_2|x)} \quad \Big| - \frac{1}{P(w_2|x)}$$

$$\frac{1}{P(w_1|x)} \le \frac{1}{P(w_2|x)}$$

$\rightarrow \quad P(w_1|x) \ge P(w_2|x) \quad \rightarrow$ assumption we made $\rightarrow$ upper bound

b.) Show using this result that for the univariate probability distributions

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

the Bayes error can be upper-bounded by

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

we know Bayes' formula: $\boxed{P(w_j|x) = \frac{p(x|w_j)\,P(w_j)}{p(x)}}$ $\quad$ posterior $= \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}$

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

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

$$P(error) \leq \int \frac{2p(x|w_1)P(w_1)p(x|w_2)\cdot P(w_2)}{p(x|w_1)P(w_1)+p(x|w_2)\cdot P(w_2)}\,dx \qquad p(x|w_1) \;\&\; p(x|w_2) \text{ einsetzen}$$

$$P(error) \leq \int \frac{2\cdot\pi^{-1}P(w_1)\,\cancel{\pi^{-1}}P(w_2)}{(1+(x-\mu)^2)(1+(x+\mu)^2)\left[\cancel{\pi^{-1}}\left(\frac{P(w_1)}{1+(x-\mu)^2} + \frac{P(w_2)}{1+(x+\mu)^2}\right)\right]}\,dx$$

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

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

$P(w_1)+P(w_2) = 1$

$$= \int \frac{2\cdot\pi^{-1}P(w_1)P(w_2)}{x^2+2\mu[P(w_1)-P(w_2)]x + 1+\mu^2} \qquad \int \frac{1}{ax^2+bx+c}\,dx = \frac{2\pi}{\sqrt{4ac-b^2}} \text{ for } b^2 < 4ac$$

$a = 1$

$b = 2\mu P(w_1) - 2\mu P(w_2)$ $\qquad b^2 = 4\mu^2\underbrace{(P(w_1)-P(w_2))^2}_{\leq 1} < 4+4\mu^2 \rightarrow$ true ◡̈

$c = 1+\mu^2$

$$P(error) \leq 2\cdot\cancel{\pi^{-1}}\cdot P(w_1)P(w_2)\cdot\frac{\cancel{2\pi}}{\sqrt{4(1+\mu^2)-4\mu^2(P(w_1)-P(w_2))^2}} = \frac{2\cdot P(w_1)P(w_2)}{\sqrt{1+\mu^2(1-P(w_1)^2+2P(w_1)P(w_2)-P(w_2)^2)}}$$

$\boxed{(P(w_1)+P(w_2))^2 = 1^2} \longrightarrow P(w_1)^2 = 1-2P(w_1)P(w_2)-P(w_2)^2$

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

Proved ◡̈

c) Explain how you would estimate the error if there was no upper-bounds that are both tight and analytically integrable.
Discuss following two cases: (1) the data is low-dimensional
(2) the data is high-dimensional

(1)
For low-dim. calculation of upper bound is fairly easy
↳ consider x not as continuous variable, but discrete → so not integrate but sum

(2)
other way: (and for high.-dim. upper bound)
→ find upper-bound numerically via Chernoff Bound or the Bhattacharyya Bound

↳ for high-dim Bhattacharyya Bound might be better, because computation less expensive than the Chernoff Bound

Exercise 2.) Bayes Decision Boundaries

One might speculate that, in some cases, the generated data $p(x|w_1)$ and $p(x|w_2)$ is of no use to improve the accuracy of a classifier, in which case one should only rely on prior class probabilities $P(w_1)$ and $P(w_2)$

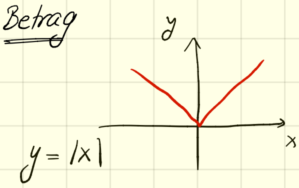
Assume: data of each class is generated by the univariate Laplacian probability distributions

$$p(x|w_1) = \frac{1}{2\theta} \exp\left(-\frac{|x-\mu|}{\theta}\right)$$
$$p(x|w_2) = \frac{1}{2\theta} \exp\left(-\frac{|x+\mu|}{\theta}\right) \qquad \mu, \theta > 0$$

a.) Compute the $\boxed{\text{Bayes optimal decision boundary} \rightarrow \text{solve } P(w_1|x) = P(w_2|x)}$

$$P(w_1|x) = P(w_2|x)$$
$$\frac{p(x|w_1)P(w_1)}{\cancel{p(x)}} = \frac{p(x|w_2)P(w_2)}{\cancel{p(x)}}$$
$$\cancel{\frac{1}{2\theta}} \exp\left(-\frac{|x-\mu|}{\theta}\right)P(w_1) = \cancel{\frac{1}{2\theta}} \exp\left(-\frac{|x+\mu|}{\theta}\right)P(w_2)$$
$$\frac{P(w_1)}{P(w_2)} = \exp\left(-\frac{|x+\mu|}{\theta} + \frac{|x-\mu|}{\theta}\right)$$
$$\theta \ln\left(\frac{P(w_1)}{P(w_2)}\right) = |x-\mu| - |x+\mu|$$

Betrag



$x \geq 0 \rightarrow |x| = x$
$x < 0 \rightarrow |x| = -x$

1. cases: $(x-\mu) < 0$ & $(x+\mu) < 0$
$$\theta \ln\left(\frac{P(w_1)}{P(w_2)}\right) = -(x-\mu) + (x+\mu) = 2\mu$$

2. case $(x-\mu) \geq 0$ & $(x+\mu) < 0$
$$\theta \ln\left(\frac{P(w_1)}{P(w_2)}\right) = (x-\mu) + (x+\mu) = 2x$$

3. Case $(x-\mu)<0$ & $(x+\mu)\geq 0$

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

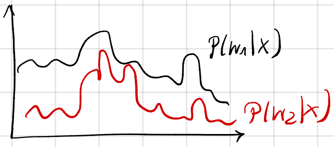
4. Case $(x-\mu)\geq 0$ & $(x+\mu)\geq 0$

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

b.) Determine for which values of $P(w_1), P(w_2), \mu, \theta$ the optimal decision is to always predict the first class (i.e. under which conditions $P(error|x) = P(w_2|x)$ $\forall x \in \mathbb{R}$

From 1a.) we know if we assume $P(w_1|x) \geq P(w_2|x)$

the $P(error|x) = P(w_2|x)$, so $P(w_1|x)$ must contain $P(w_2|x)$ →



$$P(w_1|x) \geq P(w_2|x)$$

$$\underbrace{\theta}_{\geq 0}\ \underbrace{\ln\left(\frac{P(w_1)}{P(w_2)}\right)}_{\not< 1} \geq \underbrace{|x-\mu| - |x+\mu|}_{\text{if } \mu=0}$$

$\theta \in \mathbb{R}_{\geq 0}$

$P(w_1) \geq P(w_2)$

$\mu = 0$

1. case $x<0$
$-(x)+(x)=0$

2. case $x \geq 0$
$(x)-(x)=0$

$$\Rightarrow \{(P(w_1), P(w_2), \mu, \theta) \mid P(w_1) \geq P(w_2) \wedge \mu = 0 \wedge \theta \in \mathbb{R}_{\geq 0}\}$$

c.) Repeat the exercise for the case where the data for each class is generated by the univariate Gaussian probability distributions

$$p(x|w_1) = \frac{1}{\theta\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\theta^2}\right)$$

$$p(x|w_2) = \frac{1}{\theta\sqrt{2\pi}} \exp\left(-\frac{(x+\mu)^2}{2\theta^2}\right) \qquad \theta > 0$$

a.)

optimal decision boundary

$$P(w_1|x) = P(w_2|x)$$

$$\cancel{\frac{1}{\theta\sqrt{2\pi}}} \exp\left(-\frac{(x-\mu)^2}{2\theta^2}\right) P(w_1) = \cancel{\frac{1}{\theta\sqrt{2\pi}}} \exp\left(-\frac{(x+\mu)^2}{2\theta^2}\right) P(w_2)$$

$$2\theta^2 \ln\left(\frac{P(w_1)}{P(w_2)}\right) = (x-\mu)^2 - (x+\mu)^2 = \cancel{x^2} - 2x\mu + \mu^2 - (\cancel{x^2} + 2x\mu + \mu^2) = -4x\mu$$

$$x = -\frac{1}{2} \frac{\theta^2}{\mu} \ln\left(\frac{P(w_1)}{P(w_2)}\right)$$

b.)

$$P(w_1|x) \geq P(w_2|x) \longrightarrow \underbrace{2\theta^2 \ln\left(\frac{P(w_1)}{P(w_2)}\right)}_{\theta>0 \to \ \geq 0} \geq \underbrace{-4\mu x}_{\leq 0} \longrightarrow \mu x \geq 0$$

$\mu \geq 0 \wedge x \geq 0$

$\mu < 0 \wedge$ \