Exercises for the course

Machine Learning 1

Winter semester 2024/25

Fachgebiet Maschinelles Lernen
Institut für Softwaretechnik und theoretische Informatik
Fakultät IV, Technische Universität Berlin
Prof. Dr. Klaus-Robert Müller
Email: klaus-robert.mueller@tu-berlin.de

Exercise Sheet 1

Exercise 1: Estimating the Bayes Error (10 + 10 + 10 P)

The Bayes decision rule for the two classes classification problem results in the Bayes error

$$P(\text{error}) = \int P(\text{error} \mid \boldsymbol{x})\, p(\boldsymbol{x})\, d\boldsymbol{x},$$

where $P(\text{error} \mid \boldsymbol{x}) = \min\left[P(\omega_1 \mid \boldsymbol{x}),\, P(\omega_2 \mid \boldsymbol{x})\right]$ is the probability of error for a particular input $\boldsymbol{x}$. Interestingly, while class posteriors $P(\omega_1 \mid \boldsymbol{x})$ and $P(\omega_2 \mid \boldsymbol{x})$ can often be expressed analytically and are integrable, the error function has discontinuities that prevent its analytical integration, and therefore, direct computation of the Bayes error.

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

$$P(\text{error}) \leq \int \frac{2}{\frac{1}{P(\omega_1|\boldsymbol{x})} + \frac{1}{P(\omega_2|\boldsymbol{x})}}\, p(\boldsymbol{x})\, d\boldsymbol{x}.$$

Note that the integrand is now continuous and corresponds to the harmonic mean of class posteriors weighted by $p(\boldsymbol{x})$.

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

$$p(x \mid \omega_1) = \frac{\pi^{-1}}{1 + (x-\mu)^2} \quad \text{and} \quad p(x \mid \omega_2) = \frac{\pi^{-1}}{1 + (x+\mu)^2},$$

the Bayes error can be upper-bounded by:

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

(Hint: you can use the identity $\int \frac{1}{ax^2+bx+c}\, dx = \frac{2\pi}{\sqrt{4ac-b^2}}$ for $b^2 < 4ac$.)

(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 and (2) the data is high-dimensional.

Exercise 2: Bayes Decision Boundaries (15 + 15 P)

One might speculate that, in some cases, the generated data $p(x \mid \omega_1)$ and $p(x \mid \omega_2)$ is of no use to improve the accuracy of a classifier, in which case one should only rely on prior class probabilities $P(\omega_1)$ and $P(\omega_2)$ assumed here to be strictly positive.

For the first part of this exercise, we assume that the data for each class is generated by the univariate Laplacian probability distributions:

$$p(x \mid \omega_1) = \frac{1}{2\sigma} \exp\Big(-\frac{|x-\mu|}{\sigma}\Big) \quad \text{and} \quad p(x \mid \omega_2) = \frac{1}{2\sigma} \exp\Big(-\frac{|x+\mu|}{\sigma}\Big).$$

where $\mu, \sigma > 0$.

(a) *Determine* for which values of $P(\omega_1), P(\omega_2), \mu, \sigma$ the optimal decision is to always predict the first class (i.e. under which conditions $P(\text{error} \mid x) = P(\omega_2 \mid x) \;\; \forall\, x \in \mathbb{R}$).

(b) *Repeat* the exercise for the case where the data for each class is generated by the univariate Gaussian probability distributions:

$$p(x \mid \omega_1) = \frac{1}{\sigma\sqrt{2\pi}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big) \quad \text{and} \quad p(x \mid \omega_2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\Big(-\frac{(x+\mu)^2}{2\sigma^2}\Big).$$

where $\mu, \sigma > 0$.

Exercise 3: Programming (40 P)

Download the programming files on ISIS and follow the instructions.

Exercise 1: Estimating the Bayes Error (10 + 10 + 10 P)

The Bayes decision rule for the two classes classification problem results in the Bayes error

$$P(\text{error}) = \int P(\text{error} \mid \boldsymbol{x})\, p(\boldsymbol{x})\, d\boldsymbol{x},$$

where $P(\text{error} \mid \boldsymbol{x}) = \min\left[P(\omega_1 \mid \boldsymbol{x}),\, P(\omega_2 \mid \boldsymbol{x})\right]$ is the probability of error for a particular input $\boldsymbol{x}$. Interestingly, while class posteriors $P(\omega_1 \mid \boldsymbol{x})$ and $P(\omega_2 \mid \boldsymbol{x})$ can often be expressed analytically and are integrable, the error function has discontinuities that prevent its analytical integration, and therefore, direct computation of the Bayes error.

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

$$P(\text{error}) \leq \int \frac{2}{\frac{1}{P(\omega_1|\boldsymbol{x})} + \frac{1}{P(\omega_2|\boldsymbol{x})}}\, p(\boldsymbol{x})\, d\boldsymbol{x}.$$

Note that the integrand is now continuous and corresponds to the harmonic mean of class posteriors weighted by $p(\boldsymbol{x})$.

case 1. $P(W_1|X) \leq P(W_2|X)$

$$\min[P(W_1|X), P(W_2|X)] = P(W_1|X) = \frac{2}{\frac{1}{P(W_1|X)} + \frac{1}{P(W_1|X)}} \leq \frac{2}{\frac{1}{P(W_1|X)} + \frac{1}{P(W_2|X)}}$$

case 2. $P(W_1|X) \geq P(W_2|X)$

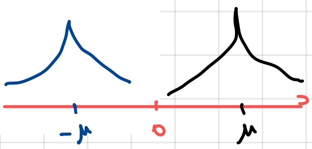
$$\min[P(W_1|X), P(W_2|X)] = P(W_2|X) = \dots \leq \dots$$

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

$$p(x \mid \omega_1) = \frac{\pi^{-1}}{1 + (x-\mu)^2} \quad \text{and} \quad p(x \mid \omega_2) = \frac{\pi^{-1}}{1 + (x+\mu)^2},$$

the Bayes error can be upper-bounded by:

Cauchy distribution



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

(Hint: you can use the identity $\int \frac{1}{ax^2+bx+c}\, dx = \frac{2\pi}{\sqrt{4ac-b^2}}$ for $b^2 < 4ac$.)

$P_1 : P(W_1)$

$P_2 : P(W_2)$

$$P(\text{error}) \leq \int \frac{2}{\frac{1}{P(W_1|X)} + \frac{1}{P(W_2|X)}}\, P(x)\, dx$$

$$= \int \frac{2}{\frac{P(x)}{P(X|W_1)P_1} + \frac{P(x)}{P(X|W_2)P_2}}\, P(x)\, dx$$

$$= \int \frac{2}{\frac{1}{P(X|W_1)P_1} + \frac{1}{P(X|W_2)P_2}}\, dx$$

$$= \int \frac{2}{\frac{1+(X-\mu)^2}{\pi^{-1} P_1} + \frac{1+(X+\mu)^2}{\pi^{-1} P_2}}\, dx$$

$$= \int \frac{2 P_1 P_2}{\pi\left[P_2(1+(X-\mu)^2) + P_1(1+(X+\mu)^2)\right]}\, dx$$

$$= 2\pi^{-1} P_1 P_2 \int \frac{1}{\underbrace{(P_1+P_2)}_{a} x^2 + \underbrace{2\mu(P_1-P_2)}_{b}X + \underbrace{1+\mu^2}_{c}}\, dx \qquad (P_1 + P_2 = 1)$$

$$b^2 < 4ac \Rightarrow 4\mu^2(P_1-P_2)^2 < 4(P_1+P_2)(1+\mu^2) \quad \checkmark$$

$$= 2\pi^{-1} P_1 P_2 \frac{2\pi}{\sqrt{4\underbrace{(P_1+P_2)}_{=1}(1+\mu^2) - 4\mu^2(P_1-P_2)^2}}$$

$$= 2\pi^{-1} P_1 P_2 \cdot \frac{2\pi}{\sqrt{4(1+\mu^2) - 4\mu^2(P_1-P_2)^2}}$$

$$= 2\pi^{-1} P_1 P_2 \frac{2\pi}{\sqrt{4 + 4\mu^2\left[1-(P_1-P_2)^2\right]}}$$

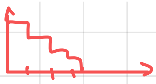
$$= 2\pi^{-1} P_1 P_2 \frac{2\pi}{\sqrt{4 + 4\mu^2\left[(P_1+P_2)^2-(P_1-P_2)^2\right]}}$$

$$= 2\pi^{-1} P_1 P_2 \frac{2\pi}{\sqrt{4 + 4\mu^2\left[(\cancel{P_1}+P_2-\cancel{P_1}+P_2)(P_1+\cancel{P_2}+P_1-\cancel{P_2})\right]}}$$

$$= P_1 P_2 \frac{2}{\sqrt{1+4\mu^2 P_1 P_2}}$$

(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 and (2) the data is high-dimensional.

(1) Riemann Sum $\int f(x)P(x)\,dx \approx \sum_i f(x_i)P(x_i)\Delta x$



(2) Monte Carlo

1. sample $x_i \sim P(x)$

2. $\int f(x)P(x)\,dx \approx \frac{1}{N}\sum f(x_i)$

Exercise 2: Bayes Decision Boundaries (15 + 15 P)

One might speculate that, in some cases, the generated data $p(x \mid \omega_1)$ and $p(x \mid \omega_2)$ is of no use to improve the accuracy of a classifier, in which case one should only rely on prior class probabilities $P(\omega_1)$ and $P(\omega_2)$ assumed here to be strictly positive.

For the first part of this exercise, we assume that the data for each class is generated by the univariate Laplacian probability distributions:

$$p(x \mid \omega_1) = \frac{1}{2\sigma}\exp\Big(-\frac{|x-\mu|}{\sigma}\Big) \quad \text{and} \quad p(x \mid \omega_2) = \frac{1}{2\sigma}\exp\Big(-\frac{|x+\mu|}{\sigma}\Big).$$

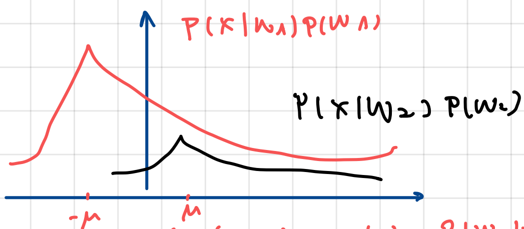
where $\mu, \sigma > 0$.

(a) *Determine* for which values of $P(\omega_1), P(\omega_2), \mu, \sigma$ the optimal decision is to always predict the first class (i.e. under which conditions $P(\text{error} \mid x) = P(\omega_2 \mid x) \quad \forall\, x \in \mathbb{R}$).

$P(W_1|X) > P(W_2|X)$

$\Leftrightarrow P(X|W_1)P(W_1) > P(X|W_2)P(W_2)$

$\Leftrightarrow \frac{\cancel{1}}{2\cancel{\delta}} e^{-\frac{|x-\mu|}{\delta}} P(W_1) > \frac{\cancel{1}}{2\cancel{\delta}} e^{-\frac{|x+\mu|}{\delta}} P(W_2)$



So that $P(\text{error}|x) = P(W_2|x)\ \forall x$

$\Leftrightarrow -\frac{|x-\mu|}{\delta} + \log P(W_1) > -\frac{|x+\mu|}{\delta} + \log P(W_2)$

① $x \leq -\mu \Rightarrow -\frac{\mu-x}{\delta} + \log P(W_1) > -\frac{-x-\mu}{\delta} + \log P(W_2)$

$\Rightarrow 2\mu < \delta(\log P(W_1) - \log P(W_2))$

(the 1. Bound)

② $-\mu < x \le \mu \Rightarrow -\frac{\mu - x}{\delta} + \log P(w_1) > -\frac{x+\mu}{\delta} + \log P(w_2)$

$$\Rightarrow \quad -2x < \delta(\log P(w_1) - \log P(w_2))$$

③ $x > \mu \Rightarrow \quad -2\mu < \delta[\log P(w_1) - \log P(w_2))$

$$\left\{ (P(w_1), P(w_2), \mu, \delta) \mid \log \frac{P(w_1)}{P(w_2)} > \frac{2\mu}{\delta} \right\}$$

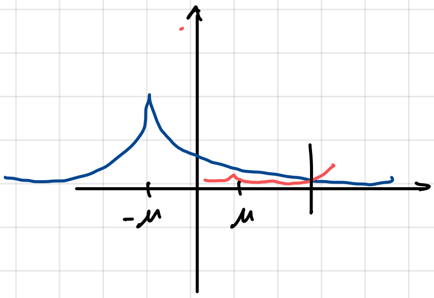
(b) *Repeat* the exercise for the case where the data for each class is generated by the univariate Gaussian probability distributions:

$$p(x \mid \omega_1) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad \text{and} \quad p(x \mid \omega_2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x+\mu)^2}{2\sigma^2}\right).$$

where $\mu, \sigma > 0$.

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

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



$$-\frac{(x-\mu)^2}{2\delta^2} + \log P(w_1) > -\frac{(x+\mu)^2}{2\delta^2} + \log P(w_2)$$

$$-2x\mu < \delta^2[\log P(w_1) - \log P(w_2)]$$

$$\Rightarrow \emptyset$$

$$\frac{1}{2}\left(-(x-\mu)^2 + (x+\mu)^2\right) > -\delta^2 \log\left(\frac{P(w_1)}{P(w_2)}\right)$$

$$\frac{1}{2}\left((x-\mu)^2 - (x+\mu)^2\right) < \delta^2 \log\left(\frac{P(w_1)}{P(w_2)}\right)$$

$$\frac{1}{2}\left((2x)(-2\mu)\right) < \delta^2 \log\left(\frac{P(w_1)}{P(w_2)}\right)$$

$$-2x\mu < \delta^2 \log\left(\frac{P(w_1)}{P(w_2)}\right)$$

2. **右边的表达式**:
$\sigma^2[\log P(\omega_1) - \log P(\omega_2)]$ 是一个固定的常数，记作 C。

3. **是否存在解?**
   - 如果 $C > 0$，即 $\log P(\omega_1) > \log P(\omega_2)$，则 $\sigma^2[\log P(\omega_1) - \log P(\omega_2)]$ 是正数。但是，由于 $-2x\mu$ 可正可负，是否存在 x 使得它严格小于 C 取决于 x 的取值范围。
   - 如果 $C < 0$，即 $P(\omega_1) < P(\omega_2)$，则 $\sigma^2[\log P(\omega_1) - \log P(\omega_2)]$ 是负数，而左侧的 $-2x\mu$ 总是可以大于这个负数，但题目要求的是 **"总是" 预测** ω_1，即不等式应当对所有 x 成立。

Exercise 1: Estimating the Bayes Error (10 + 10 + 10 P)

The Bayes decision rule for the two classes classification problem results in the Bayes error

$$P(\text{error}) = \int P(\text{error} \mid \boldsymbol{x})\, p(\boldsymbol{x})\, d\boldsymbol{x},$$

where $P(\text{error} \mid \boldsymbol{x}) = \min[P(\omega_1 \mid \boldsymbol{x}), P(\omega_2 \mid \boldsymbol{x})]$ is the probability of error for a particular input $\boldsymbol{x}$. Interestingly, while class posteriors $P(\omega_1 \mid \boldsymbol{x})$ and $P(\omega_2 \mid \boldsymbol{x})$ can often be expressed analytically and are integrable, the error function has discontinuities that prevent its analytical integration, and therefore, direct computation of the Bayes error.

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

$$P(\text{error}) \leq \int \frac{2}{\frac{1}{P(\omega_1|\boldsymbol{x})} + \frac{1}{P(\omega_2|\boldsymbol{x})}}\, p(\boldsymbol{x})\, d\boldsymbol{x}.$$

Note that the integrand is now continuous and corresponds to the harmonic mean of class posteriors weighted by $p(\boldsymbol{x})$.

a)

$$P(\text{error}) = \int P(\text{error} \mid x)\, p(x)\, dx$$

$$= \int \min[P(w_1 \mid x), P(w_2 \mid x)]\, P(x)\, dx \qquad \text{because of } \min[x_1, x_2, \dots x_n] \leq \frac{n}{\sum \frac{1}{x_i}}$$

$$= \int \frac{2}{\frac{1}{P(w_1 \mid x)} + \frac{1}{P(w_2 \mid x)}}\, P(x)\, dx$$

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

$$p(x \mid \omega_1) = \frac{\pi^{-1}}{1 + (x - \mu)^2} \quad \text{and} \quad p(x \mid \omega_2) = \frac{\pi^{-1}}{1 + (x + \mu)^2},$$

the Bayes error can be upper-bounded by:

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

(Hint: you can use the identity $\int \frac{1}{ax^2+bx+c}\, dx = \frac{2\pi}{\sqrt{4ac-b^2}}$ for $b^2 < 4ac$.)

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

$$= \int \frac{2}{\frac{1}{\frac{P(w_1 \cap x)}{P(x)}} + \frac{1}{\frac{P(w_2 \cap x)}{P(x)}}}\, P(x)\, dx$$

$$= \int \frac{2}{\frac{1}{\frac{P(x \mid w_1) P(w_1)}{\cancel{P(x)}}} + \frac{1}{\frac{P(x \mid w_2) P(w_2)}{\cancel{P(x)}}}}\, \cancel{P(x)}\, dx$$

$$= \int \frac{2}{\frac{1}{P(x \mid w_1) P(w_1)} + \frac{1}{P(x \mid w_2) P(w_2)}}\, dx$$

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

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

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

在之后使用 $P(w_1)+P(w_2)=1$ 可以减少很多计算量

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

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

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

$$4ac-b^2 = 4(P(w_2)+P(w_1))[(1+\mu^2)P(w_2)+(1+\mu^2)P(w_1)]-(-2\mu P(w_2)+2\mu P(w_1))^2$$

$$= 4(1+\mu^2)\cdot(P(w_1)+P(w_2))^2-4\mu^2(P(w_1)-P(w_2))^2$$

$$= 4(1+\mu^2)(P^2(w_1)+2P(w_1)P(w_2)+P^2(w_2))-4\mu^2(P^2(w_1)-2P(w_1)P(w_2)+P^2(w_2))$$

$$= 4P^2(w_1)+8P(w_1)P(w_2)+4P^2(w_2)+4\mu^2P^2(w_1)+8\mu^2P(w_1)P(w_2)+4\mu^2P^2(w_2)-4\mu^2(P^2(w_1)-2P(w_1)P(w_2)+P^2(w_2))$$

$$= 4P^2(w_1)+(8+16\mu^2)P(w_1)P(w_2)+4P^2(w_2) \geq 0$$

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

$$= \frac{4P(w_1)P(w_2)}{\sqrt{4P^2(w_1)+(8+16\mu^2)P(w_1)P(w_2)+4P^2(w_2)}}$$

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

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

(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 and (2) the data is high-dimensional.

For low dimensional data, it's possible to estimate the posterior probability. And we could use the optimal decision function $\arg\min_j P(w_j|x)$ to calculate the error

For high dimensional data, the error computation would be very complex. To solve this, we could reduce the dimension and map the data from a high Dimension to a low Dimension.

Exercise 2: Bayes Decision Boundaries (15 + 15 P)

One might speculate that, in some cases, the generated data $p(x \mid \omega_1)$ and $p(x \mid \omega_2)$ is of no use to improve the accuracy of a classifier, in which case one should only rely on prior class probabilities $P(\omega_1)$ and $P(\omega_2)$ assumed here to be strictly positive.

For the first part of this exercise, we assume that the data for each class is generated by the univariate Laplacian probability distributions:

$$p(x \mid \omega_1) = \frac{1}{2\sigma} \exp\Big(-\frac{|x-\mu|}{\sigma}\Big) \quad \text{and} \quad p(x \mid \omega_2) = \frac{1}{2\sigma} \exp\Big(-\frac{|x+\mu|}{\sigma}\Big).$$

where $\mu, \sigma > 0$.

(a) *Determine* for which values of $P(\omega_1), P(\omega_2), \mu, \sigma$ the optimal decision is to always predict the first class (i.e. under which conditions $P(\text{error} \mid x) = P(\omega_2 \mid x) \;\; \forall\, x \in \mathbb{R}$).

a)

$$P(W_1|x) = P(W_2|x)$$

$$\frac{P(W_1 \cdot x)}{P(x)} = \frac{P(W_2 \cdot x)}{P(x)}$$

$$\frac{P(x|W_1)P(W_1)}{\cancel{P(x)}} = \frac{P(x|W_2)P(W_2)}{\cancel{P(x)}}$$

$$\frac{\cancel{1}}{\cancel{2\delta}} e^{-\frac{|x-\mu|}{\delta}} P(W_1) = \frac{\cancel{1}}{\cancel{2\delta}} e^{-\frac{|x+\mu|}{\delta}} P(W_2)$$

$$-\frac{|x-\mu|}{\delta} + \ln P(W_1) = -\frac{|x+\mu|}{\delta} + \ln P(W_2)$$

$$\delta \ln\left(\frac{P(W_1)}{P(W_2)}\right) = |x-\mu| - |x+\mu|$$

① $x<\mu$ and $x<-\mu^{<0}$: $\quad |x-\mu| - |x+\mu| = \mu - x + x + \mu = 2\mu > 0$

② $x<\mu$ and $x \geq -\mu$: $\quad |x-\mu| - |x+\mu| = \mu - x - x - \mu = -2x \begin{cases} <0 & \text{if } 0<x<\mu \\ \geq 0 & \text{if } -\mu<x\leq 0 \end{cases}$

③ $x \geq \mu$: $(\mu>0)$ $\quad |x-\mu| - |x+\mu| = x - \mu - x - \mu = -2\mu < 0$

Decision Boundary:

$$\min(P(W_1|x), P(W_2|x)) = \begin{cases} P(W_1|x) & \text{if } x<-\mu<0 \text{ and } -\mu<x\leq 0 \;\; (x<0) \\ P(W_2|x) & \text{if } x\geq\mu \text{ and } 0<x<\mu \;\; (x \geq 0) \end{cases}$$

$$\ln\left(\frac{P(W_1)}{P(W_2)}\right) > 0 \Rightarrow P(W_1) > P(W_2) \Rightarrow P(W_1|x) - P(W_2|x) \geq 0$$

(b) *Repeat* the exercise for the case where the data for each class is generated by the univariate Gaussian probability distributions:

$$p(x \mid \omega_1) = \frac{1}{\sigma\sqrt{2\pi}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big) \quad \text{and} \quad p(x \mid \omega_2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\Big(-\frac{(x+\mu)^2}{2\sigma^2}\Big).$$

where $\mu, \sigma > 0$.