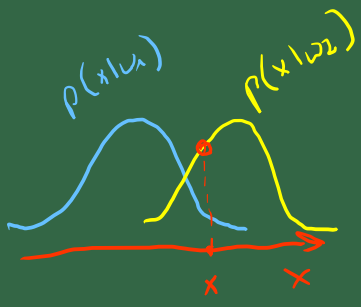


Exercise 1: Estimating the Bayes Error (10 + 10 + 10 P)

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

$$P(\text{error}) = \int P(\text{error} | \mathbf{x}) p(\mathbf{x}) d\mathbf{x},$$

where $P(\text{error} | \mathbf{x}) = \min [P(\omega_1 | \mathbf{x}), P(\omega_2 | \mathbf{x})]$ is the probability of error for a particular input \mathbf{x} . Interestingly, while class posteriors $P(\omega_1 | \mathbf{x})$ and $P(\omega_2 | \mathbf{x})$ 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:

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

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

case 1: $P(\omega_1 | x) \leq P(\omega_2 | x)$

$$\min [P(\omega_1 | x), P(\omega_2 | x)] = P(\omega_1 | x) = \frac{2}{\frac{1}{P(\omega_1 | x)} + \frac{1}{P(\omega_1 | x)}} \leq \frac{2}{\frac{1}{P(\omega_1 | x)} + \frac{1}{P(\omega_2 | x)}}$$

case 2: $P(\omega_1 | x) \geq P(\omega_2 | x)$

...

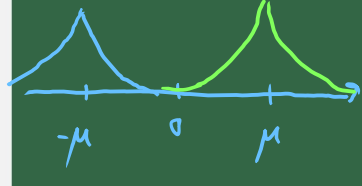
(b) Show using this result that for the univariate probability distributions

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

the Bayes error can be upper-bounded by:

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

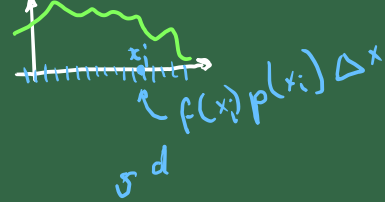
(Hint: you can use the identity $\int \frac{1}{ax^2 + bx + c} dx = \frac{2\pi}{\sqrt{4ac - b^2}}$ for $b^2 < 4ac$.)



$$\begin{aligned} P(\text{error}) &\leq \int \frac{2}{\frac{1}{P(\omega_1 | x)} + \frac{1}{P(\omega_2 | x)}} p(x) dx & p_1 &:= P(\omega_1) \\ &= \int \frac{2}{\frac{p(x)}{P(x|\omega_1)p_1} + \frac{p(x)}{P(x|\omega_2)p_2}} p(x) dx & p_2 &:= P(\omega_2) \\ &= \int \frac{2}{\frac{1+(x-\mu)^2}{\pi^{-1}p_1} + \frac{1+(x+\mu)^2}{\pi^{-1}p_2}} dx = \int \frac{2\pi^{-1}p_1p_2}{p_2(1+(x-\mu)^2) + p_1(1+(x+\mu)^2)} dx \\ &= 2\pi^{-1}p_1p_2 \cdot \int \frac{1}{\underbrace{(p_1+p_2)}_{=1}x^2 + \underbrace{2\mu(p_1-p_2)}_b \cdot x + \underbrace{1+\mu^2}_c} dx & (2\mu(p_1-p_2))^2 &< 4 \cdot (1+\mu^2) \checkmark \\ &= 4p_1p_2 \frac{1}{\sqrt{4(1+\mu^2) - (2\mu(p_1-p_2))^2}} = 4p_1p_2 \frac{1}{\sqrt{4 + 4\mu^2(p_1+p_2)^2 - 4\mu^2(p_1-p_2)^2}} \\ & \quad (p_1+p_2)^2 - (p_1-p_2)^2 = 4p_1p_2 \\ &= 4p_1p_2 \cdot \frac{1}{\sqrt{4 + 16\mu^2p_1p_2}} = \frac{2p_1p_2}{\sqrt{1 + 4\mu^2p_1p_2}} \end{aligned}$$

(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. $\int f(x)p(x)dx \approx \frac{1}{N} \sum_i f(x_i)$

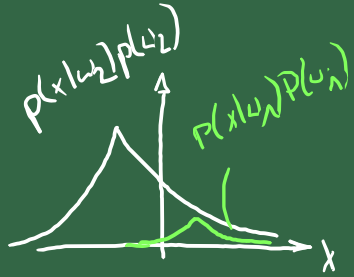
Exercise 2: Bayes Decision Boundaries (15 + 15 P)

One might speculate that, in some cases, the generated data $p(x | \omega_1)$ and $p(x | \omega_2)$ 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:

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

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 $P(\text{error} | x) = P(\omega_2 | x) \quad \forall x \in \mathbb{R}$).

$$\begin{aligned} p(\omega_1 | x) > p(\omega_2 | x) &\Leftrightarrow p(x|\omega_1)P(\omega_1) > p(x|\omega_2)P(\omega_2) \\ \Leftrightarrow \frac{1}{2\sigma} \exp\left(-\frac{|x-\mu|}{\sigma}\right)P(\omega_1) &> \frac{1}{2\sigma} \exp\left(-\frac{|x+\mu|}{\sigma}\right)P(\omega_2) \quad | \cdot 2\sigma \quad | \log \\ \Leftrightarrow -\frac{|x-\mu|}{\sigma} + \log P(\omega_1) &> -\frac{|x+\mu|}{\sigma} + \log P(\omega_2) \end{aligned}$$

① $x \leq -\mu \Rightarrow |x-\mu| = \mu-x \quad \text{and} \quad |x+\mu| = -x-\mu$

$$\Rightarrow -\frac{\mu-x}{\sigma} + \log P(\omega_1) > -\frac{-x-\mu}{\sigma} + \log P(\omega_2)$$

$$\Leftrightarrow 2\mu < \sigma \cdot [\log P(\omega_1) - \log P(\omega_2)]$$

$\left(\begin{array}{l} \textcircled{2} -\mu < x < \mu \Rightarrow -2x < \sigma [\log P(\omega_1) - \log P(\omega_2)] \\ \textcircled{3} x > \mu \Rightarrow -2\mu < \sigma [\log P(\omega_1) - \log P(\omega_2)] \end{array} \right) \text{ covered by } \textcircled{1}$

$$\left\{ (P(\omega_1), P(\omega_2), \mu, \sigma) \mid \log \frac{P(\omega_1)}{P(\omega_2)} > \frac{2\mu}{\sigma} \right\}$$

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

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

where $\mu, \sigma > 0$.

$$\begin{aligned} -\frac{(x-\mu)^2}{2\sigma^2} + \log P(\omega_1) &> -\frac{(x+\mu)^2}{2\sigma^2} + \log P(\omega_2) \\ -2x\mu &< \sigma^2 [\log P(\omega_1) - \log P(\omega_2)] \\ \Rightarrow \emptyset \end{aligned}$$

