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 $p(\boldsymbol{x})$.

Solutions:

$\because$ $P(\text{error}|x)$, $P(w_1|x)$ and $P(w_2|x)$ are all ≥ 0

$\therefore$ we only need to show:

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

$\because$ $P(\text{error}|x) = \min[P(w_1|x), P(w_2|x)]$

$\therefore$

① we first suppose that $P(w_1|x) \leq P(w_2|x)$ (some part of the density function)

$\therefore$ $P(\text{error}|x) = P(w_1|x)$

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

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

② We can also suppose that $P(w_2|x) \geq P(w_1|x)$

日期: /

then following the same proof procedure:

we can also proof that:

$$p(\omega_2|x) \le \frac{2}{\frac{1}{p(\omega_1|x)} + \frac{1}{p(\omega_2|x)}}$$

$\therefore$ Since in both 2 cases $P(\text{error}|x) \le \dfrac{2}{\frac{1}{p(\omega_1|x)} + \frac{1}{p(\omega_2|x)}}$

$\therefore$ proofed.

(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:

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

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

Solution:

$$p(\text{error}) \le \int \frac{2}{\frac{1}{p(\omega_1|x)} + \frac{1}{p(\omega_2|x)}}\, p(x)\, dx$$

$$= \int \frac{2 \times 1}{\left(\frac{1}{p(\omega_1|x)} + \frac{1}{p(\omega_2|x)}\right)\frac{1}{p(x)}}\, dx$$

$$= \int \frac{2}{\frac{1}{p(x|\omega_1)p(\omega_1)} + \frac{1}{p(x|\omega_2)p(\omega_2)}}\, dx$$

$$= \int \frac{2}{\frac{1+(x-\mu)^2}{\pi^{-1}p(\omega_1)} + \frac{1+(x+\mu)^2}{\pi^{-1}p(\omega_2)}}\, dx$$

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

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

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

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

$b^2 < 4ac$

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

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

since $(p(w_1)+p(w_2)) = 1$

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

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

proofed.

(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.

Solution:

(1) Low dimensional

we can accumulate the error using density diagram.

日期： /

1-dim/2-dim/3-dim

$$\int f(x)p(x)\,dx \xrightarrow{\text{Discretize}} \sum_i f(x_i)p(x_i) = P(\text{error})$$

(2) High Dimensional Data:

We can't integrate the density function in every dimension so we can take multiple samples, then count the number of all wrong classifications and calculate the percentage.

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

Solution:

$$P(\omega_1|x) = \frac{P(x|\omega_1)P(\omega_1)}{P(x)} \qquad P(\omega_2|x) = \frac{P(x|\omega_2)P(\omega_2)}{P(x)}$$

$\Downarrow$ take log, slides 10/25 $\qquad$ $\Downarrow$ take log, slides 10/25

$$\log P(\omega_1|x) = \log \frac{P(x|\omega_1)P(\omega_1)}{P(x)} \qquad \log P(\omega_2|x) = \log \frac{P(x|\omega_2)P(\omega_2)}{P(x)}$$

$\Downarrow$ remove all constant value $\Downarrow$

$$\log P(x|\omega_1)P(\omega_1) \qquad \log P(x|\omega_2)P(\omega_2)$$

$$\| \qquad \|$$

$$\log P(x|\omega_1) + \log P(\omega_1) \qquad \log P(x|\omega_2) + \log P(\omega_2)$$

$$\| \qquad \|$$

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

$\therefore$ If class 1 dominate the classifier, then we should have.

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

$\therefore$ according to the value of x we can remove the absolute sign by expanding the following cases.

① $x \geq \mu$

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

日期:

$$\therefore \frac{2\mu}{\sigma} > \log\frac{P(w_2)}{P(w_1)} \Longrightarrow -\frac{2\mu}{\sigma} < \log\frac{P(w_1)}{P(w_2)}$$

$\therefore$ the relationship of $\mu, \sigma, P(w_2), P(w_1)$ when $x \geq \mu$ is shown above.

② $-\mu < x < \mu$

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

$$\therefore \frac{2x}{\sigma} > \log\frac{P(w_2)}{P(w_1)} \longrightarrow -\frac{2x}{\sigma} < \log\frac{P(w_1)}{P(w_2)}$$

$$\Downarrow$$

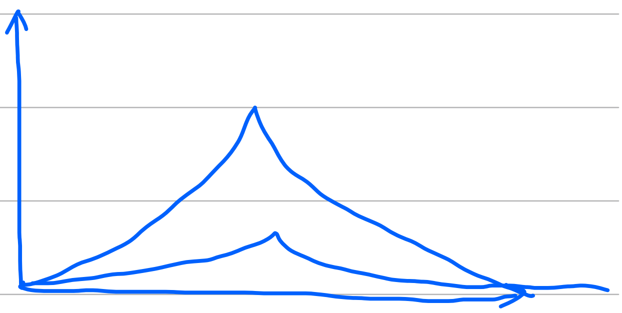
$$\frac{2\mu}{\sigma} < \log\frac{P(w_1)}{P(w_2)}$$

③ $x \leq -\mu$

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

$$\therefore -\frac{2\mu}{\sigma} > \log\frac{P(w_2)}{P(w_1)} \Longrightarrow \frac{2\mu}{\sigma} < \log\frac{P(w_1)}{P(w_2)}$$

$$\left.\begin{cases} -\dfrac{2\mu}{\sigma} < \log\left(\dfrac{P(w_1)}{P(w_2)}\right) \\ \dfrac{2\mu}{\sigma} < \log\left(\dfrac{P(w_1)}{P(w_2)}\right) \\ \dfrac{2\mu}{\sigma} < \log\left(\dfrac{P(w_1)}{P(w_2)}\right) \end{cases}\right\} \text{combine} \longrightarrow \frac{2\mu}{\sigma} < \log\frac{P(w_1)}{P(w_2)}$$



(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 $\mu, \sigma > 0$.

Solution:

$$-\frac{(x-\mu)^2}{2\sigma^2} + \log P(\omega_1) > -\frac{(x+\mu)^2}{2\sigma^2} + \log P(\omega_2)$$

$$\Downarrow$$

$$\frac{2\mu x}{2\sigma^2} + \log P(\omega_1) > -\frac{2\mu x}{2\sigma^2} + \log P(\omega_2)$$

$$\frac{2\mu x}{\sigma^2} > \log\Big(\frac{P(\omega_2)}{P(\omega_1)}\Big)$$

$$-x < \frac{\sigma^2}{2\mu} \log\Big(\frac{P(\omega_1)}{P(\omega_2)}\Big) \quad ①$$

$\therefore$ If we want to guarantee to take class one then we should try to make ① hold.