TECHNISCHE UNIVERSITÄT BERLIN
INSTITUTE OF GEODESY AND GEOINFORMATION SCIENCE

Homework #1 Bayes Estimation

Student name: *Bing Li*

Course: *Machine Learning* – Professor: *Prof. Dr. Klaus-Robert Mueller*
Due date: *October 21th, 2019*

Question 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 $P(error) = \int P(error|x)p(x)dx$. where $P(error|x) = min[P(\omega 1|x), P(\omega 2|x)]$is the probability of error for a particular input x. Interestingly, while class posteriors $P(\omega 1|x) and P(\omega 2|x)$ 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:

$$P(error) \leq \int \frac{2}{\frac{1}{P(\omega 1|x)} + \frac{1}{P(\omega 2|x)}} p(x)dx \tag{1}$$

Answer. Proof:
$\because P(error|x) = min[P(\omega 1|x), P(\omega 2|x)] \implies \begin{cases} P(\omega 1|x) \geq P(error|x) \\ P(\omega 2|x) \geq P(error|x) \end{cases}$
$\therefore \frac{2}{\frac{1}{P(\omega 1|x)} + \frac{1}{P(\omega 2|x)}} \geq \frac{2}{\frac{1}{P(error|x)} + \frac{1}{P(error|x)}} = P(error|x)$
$\therefore \frac{2}{\frac{1}{P(\omega 1|x)} + \frac{1}{P(\omega 2|x)}} p(x) \geq P(error|x)p(x)$
$\therefore P(error) \leq \int \frac{2}{\frac{1}{P(\omega 1|x)} + \frac{1}{P(\omega 2|x)}} p(x)dx$

(b) Show using this result that for the univariate probability distributions $p(x|\omega_1) = \frac{\pi^{-1}}{1+(x-\mu)^2}$ and $p(x|\omega_2) = \frac{\pi^{-1}}{1+(x+\mu)^2}$
the Bayes error can be upper-bounded by: $P(error) \leq \frac{2P(\omega_1)P(\omega_2)}{\sqrt{1+4\mu^2 P(\omega_1)P(\omega_2)}}$

Answer. $P(\omega_1|x) = \frac{P(\omega_1)p(x|\omega_1)}{p(x)}$
$P(\omega_2|x) = \frac{P(\omega_2)p(x|\omega_2)}{p(x)}$
insert the above into (a),then: $P(error) \leq \int \frac{2P(\omega_1)P(\omega_2)}{\frac{P(\omega_2)}{p(x|\omega_1)} + \frac{P(\omega_1)}{p(x|\omega_2)}} p(x)dx$

then insert $p(x|\omega_1) = \frac{\pi^{-1}}{1+(x-\mu)^2}$ and $p(x|\omega_2) = \frac{\pi^{-1}}{1+(x+\mu)^2}$ into the above, we get:

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

then we use $\int \frac{1}{ax^2+bx+c}dx = \frac{2\pi}{\sqrt{4ac-b^2}}$

$\therefore \int \frac{2P(\omega_1)P(\omega_2)}{\pi[(P(\omega_1)+P(\omega_2)x^2+2\mu((P(\omega_1)-P(\omega_2))x+(1+\mu^2)(P(\omega_1)+P(\omega_2))]} dx =$
$\frac{2P(\omega_1)P(\omega_2)}{\pi} \int \frac{1}{(P(\omega_1)+P(\omega_2)x^2+2\mu((P(\omega_1)-P(\omega_2))x+(1+\mu^2)(P(\omega_1)+P(\omega_2))} dx =$
$\frac{2P(\omega_1)P(\omega_2)}{\pi} \frac{2\pi}{\sqrt{4(\mu^2+1)(P(\omega_1)+P(\omega_2))^2-4\mu^2(P(\omega_1)-P(\omega_2))^2}} = \frac{2P(\omega_1)P(\omega_2)}{\sqrt{1+4\mu^2P(\omega_1)P(\omega_2)}}$

so finally we get the conclusion

(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.

Answer. if the data is low-dimensional, we can assume the parameter of the conditional densities and use the maximum likelihood estimation to calculate the parameters. If the data is high-dimensional,we can use Bayesian estimation to predict the parameter of distribution,the parameters follow a conditional distribution.

ex.2.

(a).

According to Bayes rules: $P(w_i|x) = \frac{P(x|w_i)P(w_i)}{P(x)}$

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

$\because$ Bayes decision rule: $\begin{cases} w_1 & \text{if } P(w_1|x) > P(w_2|x) \\ w_2 & \text{else.} \end{cases}$

if we want to always predict the first class.

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

$$\Rightarrow P(w_1|x) = \frac{P(x|w_1)\cdot P(w_1)}{P(x)} > P(w_2|x) = \frac{P(x|w_2)P(w_2)}{P(x)}$$

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

$$\frac{1}{2\sigma}\cdot e^{-\frac{|x-\mu|}{\sigma}}\cdot P(w_1) > \frac{1}{2\sigma}\cdot e^{-\frac{|x+\mu|}{\sigma}}\cdot P(w_2)$$

$\because$ $\mu, \sigma > 0$ $\quad P(w_1), P(w_2) \geq 0$

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

for $f(x) = e^{\frac{|x+\mu|-|x-\mu|}{\sigma}} = \begin{cases} e^{\frac{-2\mu}{\sigma}} & x \leq -\mu \\ e^{\frac{2x}{\sigma}} & -\mu < x < \mu \\ e^{\frac{2\mu}{\sigma}} & x \geq \mu \end{cases}$

$$\therefore \frac{P(w_2)}{P(w_1)} < e^{-\frac{2\mu}{\sigma}}$$

(b).

$\because \; P(x|w_1)\cdot P(w_1) > P(x|w_2)\cdot P(w_2)$

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

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

$$e^{\frac{2x\cdot\mu}{\sigma^2}} > \frac{P(w_2)}{P(w_1)}$$

make $\mu=0$ than $e^{\frac{2\cdot x\cdot\mu}{\sigma^2}} = 1$

than $\frac{P(w_2)}{P(w_1)} < 1$ and $\mu = 0$