Machine Learning I WS 2018/2019

Group name : XTBQJ

Exercise Sheet 1

Exercise 1 :

(a) with $P(\text{error}) = \int p(\text{error}|x) \cdot p(x) \cdot dx$

and we want to show that $\int p(\text{error}|x) \cdot p(x) \cdot dx \leq \int \frac{2}{\frac{1}{p(w_1|x)} + \frac{1}{p(w_2|x)}} \cdot p(x) \cdot dx$

i.e $\min[p(w_1|x), p(w_2|x)] \leq \frac{2}{\frac{1}{p(w_1|x)} + \frac{1}{p(w_2|x)}}$

must hold for $\forall x$.

as we know that $p(w_1|x) + p(w_2|x) = 1$, so $\min[p(w_1|x), p(w_2|x)] \leq 0.5$ will always be true

and $\max[p(w_1|x), p(w_2|x)] \geq 0.5$ will also be true,

then $2 \cdot \max[p(w_1|x), p(w_2|x)] \geq 1$, with $\min[p(w_1|x), p(w_2|x)] \geq 0$.

$\Rightarrow \min[p(w_1|x), p(w_2|x)] \leq 2 \cdot \max[p(w_1|x), p(w_2|x)] \cdot \min[p(w_1|x), p(w_2|x)]$

$\Rightarrow$ ~~Assume that~~ $\max[p(w_1|x), p(w_2|x)] \cdot \min[p(w_1|x), p(w_2|x)] == p(w_1|x) \cdot p(w_2|x)$.

$\Rightarrow \min[p(w_1|x), p(w_2|x)] \leq 2 \cdot p(w_1|x) \cdot p(w_2|x)$

$\Rightarrow \min[p(w_1|x), p(w_2|x)] \leq \frac{2 \cdot p(w_1|x) \cdot p(w_2|x)}{p(w_1|x) + p(w_2|x)}$ with $p(w_1|x) + p(w_2|x) = 1$

$\Rightarrow \min[p(w_1|x), p(w_2|x)] \leq \frac{2}{\frac{1}{p(w_1|x)} + \frac{1}{p(w_2|x)}}$

i.e $p(\text{error}|x) \leq \frac{2}{\frac{1}{p(w_1|x)} + \frac{1}{p(w_2|x)}}$

So we could proof that the full error $P(\text{error})$ has an upper-bound.

(b).

with the result above, we have shown that:

$$P(\text{error}) \leq \int \frac{2}{\frac{1}{p(w_1|x)} + \frac{1}{p(w_2|x)}} \cdot p(x) \cdot dx$$, and with Bayes-theory.

It could be simplified as :

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

$$\Rightarrow \int \frac{2 p(x|w_1) \cdot p(x|w_2) \cdot p(w_1) \cdot p(w_2)}{p(x|w_2) \cdot p(w_2) + p(x|w_1) \cdot p(w_1)} dx \Rightarrow 2 p(w_1) p(w_2) \cdot \int \frac{p(x|w_1) \cdot p(x|w_2)}{p(x|w_2) \cdot p(w_2) + p(x|w_1) \cdot p(w_1)} dx$$

and with the given ~~of~~ probability distributions of $p(x|w_1)$ and $p(x|w_2)$.

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

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

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

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

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

in this case, $b^2=4\mu^2(p(w_1)-p(w_2))^2$. $4ac=4\mu^2+4$

with ~~xxx~~ $p(w_1)-p(w_2)<1$, $b^2<4ac$ will always be true.

so the integral could be written as:

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

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

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

$$\Rightarrow 2\cdot p(w_1)\cdot p(w_2)\cdot\frac{1}{\sqrt{1+\mu^2\cdot 2p(w_1)\cdot 2\cdot p(w_2)}} \quad \text{with } 1=p(w_1)+p(w_2)$$

$$\Rightarrow\frac{2\cdot p(w_1)\cdot p(w_2)}{\sqrt{1+4\mu^2 p(w_1)\cdot p(w_2)}}$$

thus we could find a upper-bound of P(error) for the given distribution.

(c)

(1) For low-dimensional data, the numerical integration can be used to get a approximate solution as upper-bound of error.

(2) For high-dimensional data, numerical integration would takes much more complicated calculation, in this case, Monte-Carlo Method is introduced for Integration of high dimension.

Exercise 2:

(a) According to Bayes decision, the optimal decision boundary is: $p(w_1|x)=p(w_2|x)$

with Bayes theory, it could be written as: $p(x|w_1)\cdot p(w_1)=p(x|w_2)\cdot p(w_2)$

take ~~to~~ ~~at~~ logarithm on both sides: $\ln p(x|w_1)+\ln p(w_1)=\ln p(x|w_2)+\ln p(w_2)$

with the given probability distributions:

$\Rightarrow -\log 2\sigma-\frac{|x-\mu|}{\sigma}+\ln p(w_1)=-\log 2\sigma-\frac{|x+\mu|}{\sigma}+\ln p(w_2)$

$\Rightarrow |x-\mu|-|x+\mu|=\sigma\cdot\ln\frac{p(w_1)}{p(w_2)}$

so the boundary will be $x+b_0=0$, with $b_0=-\frac{\sigma}{2}\ln\frac{p(w_1)}{p(w_2)}$

(b) If the optimal decision always predicts the first class, then $p(w_1|x)>p(w_2|x)$ must holds for every x.

$\Rightarrow -\frac{|x-\mu|}{\sigma}+\ln p(w_1)>-\frac{|x+\mu|}{\sigma}+\ln p(w_2)$ must always be true

$\Rightarrow |x+\mu|-|x-\mu|>-\sigma\cdot\ln\frac{p(w_1)}{p(w_2)}$ with $\mu,\sigma>0$.

$|x+\mu|-|x-\mu|\in[-2\mu,2\mu]$

So let $-2\mu>-\sigma\cdot\ln\frac{p(w_1)}{p(w_2)}\Rightarrow\frac{2\mu}{\sigma}<\ln\frac{p(w_1)}{p(w_2)}$

i.e when $\frac{p(w_1)}{p(w_2)}>e^{\frac{2\mu}{\sigma}}$, the decision will always predict class w_1.

(c) The optimal decision boundary will still be $p(w_1|x)=p(w_2|x)$,

and with Bayes theory: $p(x|w_1)\cdot p(w_1)=p(x|w_2)\cdot p(w_2)$

$\Rightarrow \ln p(x|w_1)+\ln p(w_1)=\ln p(x|w_2)+\ln p(w_2)$

$\Rightarrow -\ln\sigma-\ln\sqrt{2\pi}-\frac{(x-\mu)^2}{2\sigma^2}+\ln p(w_1)=-\ln\sigma-\ln\sqrt{2\pi}-\frac{(x+\mu)^2}{2\sigma^2}+\ln p(w_2)$

$\Rightarrow \cancel{+}\ (x-\mu)^2-(x+\mu)^2=2\sigma^2\ln\frac{p(w_1)}{p(w_2)}$

$\Rightarrow x\cdot\mu=-\frac{\sigma^2}{2}\ln\frac{p(w_1)}{p(w_2)}$

So the decision boundary will be: $w\cdot x+b=0$, with $w=\mu$, $b=\frac{\sigma^2}{2}\ln\frac{p(w_1)}{p(w_2)}$

And if the decision always predict w_1, then $p(w_1|x)>p(w_2|x)$ must always holds.

So $-\frac{(x-\mu)^2}{2\sigma^2}+\ln p(w_1)>-\frac{(x+\mu)^2}{2\sigma^2}+\ln p(w_2)$

$\Rightarrow (x-\mu)^2-(x+\mu)^2<2\sigma^2\ln\frac{p(w_1)}{p(w_2)}$ with $\sigma>0$

$\Rightarrow \mu x>-\frac{\sigma^2}{2}\ln\frac{p(w_1)}{p(w_2)}$.

i.e $\frac{p(w_1)}{p(w_2)}>e^{-\frac{2\mu}{\sigma^2}x}$ must always be true if the decision always predict w_1.