HW2 吴瑞林 电计1701-201785072

Problem 1

Bag A: 50 black, 100 white

Bag B: 30 white, 90 black

(a) if event A means 2b, 1w,

$$P(A) = \left(\frac{50}{150}\right)^2 \cdot \left(\frac{100}{150}\right) + \frac{50}{150} \cdot \frac{100}{150} \cdot \frac{50}{150} + \frac{100}{150} \cdot \left(\frac{50}{150}\right)^2 = \frac{2}{9}$$

(b) event B: 2b, 1w

$$P(B) = C_3^1 \times \frac{30}{120} \times \left(\frac{90}{120}\right)^2 = \frac{27}{64}$$

(c) event C: 2b, 1w, from A

$$P(C) = \frac{\frac{1}{2}P(A)}{\frac{1}{2}P(A)+\frac{1}{2}P(B)} = \frac{\frac{2}{9}}{\frac{2}{9}+\frac{27}{64}} = \frac{128}{371}$$

Problem 2

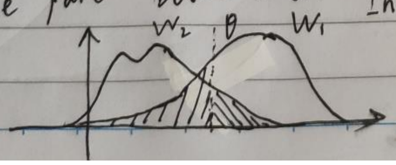
(a) Decision: $x > \theta ? w_1 : w_2$

$$P(error) = P(w_1)\int_{-\infty}^{\theta} p(x|w_1)\,dx + P(w_2)\int_{\theta}^{\infty} p(x|w_2)\,dx$$

when $x<\theta$, $P(error) = P(x|w_1)P(w_1)$

$x>\theta$, $P(error) = P(x|w_2)P(w_2)$

It means the part below in G1 In "///" part, means

G1: 

Date

a point belongs to W_1 but decided to W_2,
The area of "////" shows the error of this situation.

In "\\\\" pare, means a point belongs to W_2 but decided to W_1, the area of "\\\\" shows the error area.

(b) Rule: $P(w_1|X) > P(w_2|x)$? $w_1 : w_2$

$p(error) = 1 - \int P(W_{max}|X)\, p(x)\, dx$

where $P(W_{max}|X) \geq P(w_i|X)$ for all $i, i=1,2$

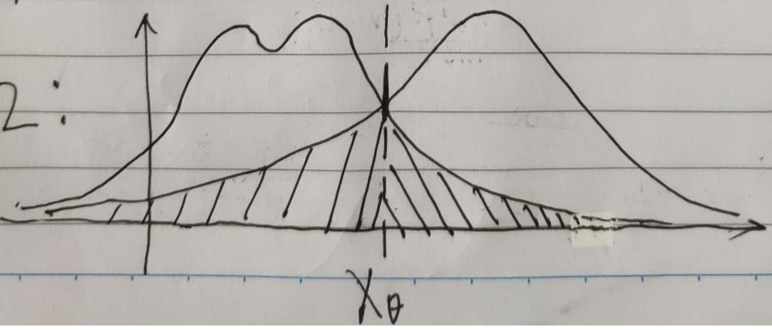
Because W_1 and W_2 are from $-\infty$ to $+\infty$,
all points are in W_1 or W_2,

$\therefore \quad P(w_1|X) + P(W_2|X) = 1$

$\therefore \quad P(W_{min}|X) + P(W_{max}|X) = 1$

$P(W_{min}|X)$ means the smallest one, when $W_1 > W_2$
$P(W_{min}|X) = P(W_2|x)$, else means $P(w_1|X)$, ~~so~~ the graph below shows $P(W_{min}|X)$ ("////" and "\\\\")

G2:



When occur $P(W_{min}|X)$, points below to W_2 are decided to W_1 and points below to W_1 are decided to W_2.

$\therefore P(error) = \int_{-\infty}^{+\infty} P(W_{min}|X) \cdot p(X)\, dx$

$$= \int_{-\infty}^{+\infty} (1 - P(W_{max}|X)) \cdot p(x)\, dx$$

$$= 1 - \int_{-\infty}^{+\infty} P(W_{max}|X)\, p(x)\, dx$$

(c) In graph 2, there is X_θ, when $\theta = X_\theta$, the two decision rules are equivalent

(d) $P(w_i|X) = \frac{P(X|w_i) \cdot P(w_i)}{P(X)}$ (Bayes Rule)

$P(X) = \sum_{j=1}^{m} P(X|w_j)\, P(w_j)$

obviously, $P(X|w_i)P(w_i)$ is part of $P(X)$

so define $x_i = P(X|w_i)\,P(w_i)$

$p(X) = x_1 + \cdots + X_m$

$$P(W_{max}|X) = \frac{P(X|W_{max})P(W_{max})}{P(X)}$$

$$= \frac{X_{max}}{P(X)} = \frac{X_{max}}{X_1 + \cdots + X_m} = \frac{1}{X_1 + \cdots + X_m} \frac{m X_{max}}{m}$$

$$\because \frac{1}{m} = \frac{X_1 + \cdots + X_m}{X_1 + \cdots + X_m} \frac{1}{m} = \frac{1}{X_1 + \cdots + X_m} \cdot \frac{X_1 + \cdots + X_m}{m}$$

$$P(W_{max}|X) - \frac{1}{m} = \frac{1}{x_1+\cdots+x_m}\frac{m x_{max}}{m} - \frac{1}{x_1+\cdots+x_m}\frac{x_1+\cdots+x_m}{m}$$

$$= \frac{1}{x_1+\cdots+x_m}\cdot\frac{1}{m}\cdot\left(\sum_{i=1}^{m} x_{max} - x_i\right)$$

$\because x_{max} \geq x_i$

$\therefore x_{max} - x_i \geq 0$

$\therefore P(W_{max}|X) - \frac{1}{m} \geq 0 \Rightarrow P(W_{max}|X) \geq \frac{1}{m}$

$$P(error) = 1 - \int P(W_{max}|X)\, p(x)\, dx$$

$$= 1 - \sum_{i=1}^{m}\int P_i(W_{max}|X)\cdot p(x)\, dx$$

define $P_i(W_{max}|X)$, means when is pare i, $P_i(W_i|X)$ is the max, the subsection defined as i. 由规则定义每个W_i在其取到 max 的区域为$P_i(W_i|X)$

$\therefore P(error) = 1 - P(W_{max}|X)\sum_{i=1}^{m}\int P_i(x)\, dx$

$\leq 1 - \frac{1}{m}\sum_{i=1}^{m}\int P_i(x)\, dx$

$\because \sum_{i=1}^{m}\int P_i(x)\, dx = 1$

$\therefore P(error) \leq 1 - \frac{1}{m} = \frac{m-1}{m}$

$\therefore P(error) \leq \frac{m-1}{m}$

problem 3:

(a) $\cancel{P(D|\theta)=}\ \theta = P(w_i)$

$P(D|\theta) = P(z_{i1}, \dots, z_{in}|\theta)$

(b) $P(D|\theta) = P(z_{i1}, \dots, z_{in}|\theta) = \prod_{k=1}^{n} P(z_{ik}|\theta)$

$$P(z_{ik}|\theta) = \begin{cases} \theta, & \text{when } z_{ik}=1 \\ 1-\theta, & \text{when } z_{ik}=0 \end{cases}$$

when $z_{ik}=1$, $x_k = z_{ik} = 1$, in $\theta^{x_k}(1-\theta)^{1-x_k}$,

$\theta^{x_k} = \theta$, $(1-\theta)^{1-x_k} = (1-\theta)^{1-1} = (1-\theta)^0 = 1$

$\therefore$ when $z_{ik}=1$, $\theta^{x_k}(1-\theta)^{1-x_k} = \theta$

$\therefore$ when $z_{ik}=0$, $\theta^{x_k}(1-\theta)^{1-x_k} = 1-\theta$

$\therefore$ $P(z_{ik}|\theta) = \theta^{x_k}(1-\theta)^{1-x_k}$

$\therefore$ $P(D|\theta) = \prod_{k=1}^{n} \theta^{x_k}(1-\theta)^{1-x_k}$, $(x_k = z_{ik})$

(c) $\ln P(D|\theta) = \sum_{k=1}^{n} \ln(\theta^{x_k}(1-\theta)^{1-x_k}) = \sum_{k=1}^{n}(x_k \ln\theta + (1-x_k)\ln(1-\theta))$

$\hat{\theta} = \operatorname{argmax}_\theta (L(\theta|D)) = \operatorname{argmax}_\theta (P(D|\theta))$

$$\frac{\partial \ln P(D|\theta)}{\partial\theta} = \sum_{k=1}^{n}\left(x_k \frac{1}{\theta} - (1-x_k)\frac{1}{1-\theta}\right)$$

$$= \sum_{k=1}^{n}\left(x_k \frac{1}{\theta} + (x_k-1)\frac{1}{1-\theta}\right) = 0$$

$\therefore \sum_{k=1}^{n}\left(x_k \frac{(1-\theta)}{\theta(1-\theta)} + x_k\frac{\theta}{(1-\theta)\theta} - \frac{\theta}{(1-\theta)\theta}\right) = 0$

$\sum_{k=1}^{n}\left(\frac{x_k-\theta}{\theta(1-\theta)}\right) = 0$

Date

when $\sum_n \frac{x_k-\theta}{\theta(1-\theta)}$ $\because (1-\theta)\theta$ has no k, $(1-\theta)\theta \neq 0$

$\therefore \sum_{k=1}^{n}(X_k-\theta)=0$, $\sum_{k=1}^{n} X_k = \sum_{k=1}^{n}\theta$

$\therefore n\theta = \sum_{k=1}^{n} X_k$, $\hat{\theta} = \frac{1}{n}\sum_{k=1}^{n} X_k$

(d) $p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} = \frac{p(D|\theta)p(\theta)}{\int p(D|\theta)p(\theta)d\theta}$

$\because p(\theta) \sim U(0,1) = \begin{cases} 1, & 0<\theta<1 \\ 0, & \text{otherwise.} \end{cases}$

$\therefore p(\theta|D) = \frac{p(D|\theta)p(\theta)}{\int_0^1 p(D|\theta)d\theta}$

$\because p(D|\theta) = \prod_{k=1}^{n} p(X_k|\theta)$ (in (a), (b))

$= \prod_{k=1}^{n} \theta^{X_k}(1-\theta)^{1-X_k}$

$\therefore p(\theta|D) = \frac{\prod_{k=1}^{n}\theta^{X_k}(1-\theta)^{1-X_k}}{\int_0^1 \prod_{k=1}^{n}\theta^{X_k}(1-\theta)^{1-X_k}d\theta}$ $(0<\theta<1)$

$\because \int_0^1 \theta^m(1-\theta)^n d\theta = \frac{m!\,n!}{(m+n+1)!}$, $m=\prod_{k=1}^{n} X_k$, $n=\prod_{k=1}^{n} 1-X_k$

$\therefore p(\theta|D) = \frac{\prod_{k=1}^{n}\theta^{X_k}(1-\theta)^{1-X_k}}{\frac{(\prod_{k=1}^{n} X_k)!\cdot(\prod_{k=1}^{n} 1-X_k)!}{(n+1)!}}$

$\therefore p(\theta|D) = \begin{cases} \frac{(n+1)!\,\theta^{\sum_{k=1}^{n} X_k}(1-\theta)^{n-\sum_{k=1}^{n} X_k}}{(\prod_{k=1}^{n} X_k)!\,(\prod_{k=1}^{n} 1-X_k)!}, & 0<\theta<1 \\ 0, & \text{otherwise} \end{cases}$

(e) $p(X|D)=\int p(X,\theta|D)\,d\theta=\int p(X|\theta)\cdot p(\theta|D)\,d\theta$

$=\int_0^1 \theta^X(1-\theta)^{1-X}\cdot\frac{(n+1)!\,\prod_{k=1}^{n}\theta^{X_k}(1-\theta)^{1-X_k}}{(\sum_{k=1}^{n}X_k)!\,(\sum_{k=1}^{n}1-X_k)!}\,d\theta$

$=\frac{(n+1)!}{(\sum_{k=1}^{n}X_k)!\,(\sum_{k=1}^{n}(1-X_k))!}\int_0^1\theta^{X+\sum_{k=1}^{n}X_k}(1-\theta)^{1-X+n-\sum_{k=1}^{n}X_k}\,d\theta$

$=\frac{(n+1)!}{(\sum_{k=1}^{n}X_k)!\,(\sum_{k=1}^{n}1-X_k)!}\cdot\frac{(X+\sum_{k=1}^{n}X_k)!\,(n+1-X-\sum_{k=1}^{n}X_k)!}{(\cancel{X+\sum_{k=1}^{n}X_k}+n+1\cancel{-X-\sum_{k=1}^{n}X_k}+1)}$

$=\frac{(n+1)!}{(n+2)!}\cdot\frac{(X+\sum_{k=1}^{n}X_k)!\,(n+1-X-\sum_{k=1}^{n}X_k)!}{(\sum_{k=1}^{n}X_k)!\,(\sum_{k=1}^{n}1-X_k)!}$

$=\frac{1}{n+2}\cdot\frac{(X+\sum_{k=1}^{n}X_k)!\,(n+1-X-\sum_{k=1}^{n}X_k)!}{(\sum_{k=1}^{n}X_k)!\,(\sum_{k=1}^{n}1-X_k)!}$

(f) $X=1$, $p(X|D)=\frac{1}{n+2}\frac{(1+\sum_{k=1}^{n}X_k)!\,\cancel{(n-\sum_{k=1}^{n}X_k)!}}{(\sum_{k=1}^{n}X_k)!\,\cancel{(n-\sum_{k=1}^{n}X_k)!}}$

$=\frac{1}{n+2}(1+\sum_{k=1}^{n}X_k)$

$p(D|\theta)=p(\theta|D)p(D)/p(\theta)$, $p(\theta|D)=p(D|\theta)\,p(\theta)/p(D)$

$\hat{\theta}_{BPE}=\arg\max_\theta\frac{p(X|\theta)\,p(\theta)}{p(X)}$

$p(X)$ is not related to θ

$\hat{\theta}_{BPE}=\arg\max_\theta p(X|\theta)\,p(\theta)=\arg\max_\theta(\ln p(X|\theta)+\ln p(\theta))$

$\because p(\theta|D)=p(D|\theta)\,p(\theta)/p(D)$

$=\frac{p(D|\theta)\,p(\theta)}{\int p(D|\theta)\,p(\theta)\,d\theta}$

$\hat{\theta}=\int\theta\,p(\theta|D)\,d\theta=\int p(X=1,\theta|D)\,d\theta=p(X=1|D)$

$=\frac{1}{n+2}(1+\sum_{k=1}^{n}X_k)$

No.

Date

(g) MLE : $\hat{\theta} = \frac{1}{n}\sum_{k=1}^{n} x_k$

BPE : $\hat{\theta} = \frac{1}{n+2}\left(1+\sum_{k=1}^{n} x_k\right)$

when $n \to \infty$, $n+2 = n$, $1+\sum_{k=1}^{n} x_k = \sum_{k=1}^{n} x_k$.

$\therefore$ when have ∞ sample, MLE and BPE are the same.

When samples are limited, BPE has prior knowledge and has a correction, which can perform better then MLE