Ques :- 1 Given: $P(x|\theta) = \prod_{i=1}^{d} \theta_i^{x_i} (1-\theta_i)^{1-x_i}$ → for one point → (1)

MT19133
Kaushal Sanadhya

∴ for all points i.e $x_K \ \forall \ K \in [1, N]$
eq (1) will become

$$P(D|\theta) = \prod_{K=1}^{n} \prod_{i=1}^{d} \theta_{iK}^{x_{iK}} (1-\theta_{iK})^{1-x_{iK}}$$

taking log both the sides

$$\log P(D|\theta) = \sum_{K=1}^{n} \left(\sum_{i=1}^{d} x_{iK} \log\theta_{iK} + \sum_{i=1}^{d} (1-x_{iK}) \log (1-\theta_{iK}) \right) \rightarrow (2)$$

expanding $\sum_{i=1}^{d}$ in eq (2)

$$\log P(D|\theta) = \sum_{K=1}^{n} \Big[x_{1K} \log\theta_{1K} + x_{2K} \log\theta_{2K} + \dots x_{dK} \log\theta_{dK} + (1-x_{1K}) \log(1-\theta_{K}) + (1-x_{2K}) \log(1-\theta_{2K}) + \dots + (1-x_{dK}) \log(1-\theta_{dK}) \Big] \rightarrow (3)$$

differentiating eq^n (3) partially wrt $\frac{\partial}{\partial \theta_{iK}}$ and equating to 0 to calculate maximum likelihood estimate for θ

$$\sum_{K=1}^{n} \left[\frac{x_{1K}}{\theta_{1K}} + \frac{x_{2K}}{\theta_{2K}} + \dots \frac{x_{dK}}{\theta_{dK}} - \frac{(1-x_{1K})}{(1-\theta_{1K})} - \frac{(1-x_{2K})}{(1-\theta_{2K})} - \dots - \frac{(1-x_{dK})}{(1-\theta_{dK})} \right] \rightarrow (4)$$

we know that $X = [x_1, x_2 \dots x_d]^T$ vector ∴ eq^n (4) can be written as

$$\sum_{K=1}^{n} \left[\frac{X_K}{\theta_K} - \frac{(1-X_K)}{(1-\theta_K)} \right] = 0 \rightarrow (5)$$

Simplifying eq^n (5)

$$\sum_{K=1}^{n} \frac{X_K}{\theta_K} = \sum_{K=1}^{n} \frac{(1-X_K)}{(1-\theta_K)} \quad \cancel{\sum_{K=1}^{n} \frac{X_K}{\theta_K}} =$$

$$\Rightarrow \sum_{k=1}^{n} \frac{(1-\theta_k)}{\theta_k} = \sum_{k=1}^{n} \frac{1-x_k}{x_k}$$

$$\Rightarrow \sum_{k=1}^{n} \frac{1}{\theta_k} \cancel{-1} = \sum_{k=1}^{n} \frac{1}{x_k} \cancel{-1}$$

$$\Rightarrow \sum_{k=1}^{n} \theta_k = \sum_{k=1}^{n} x_n$$

$$\Rightarrow n\hat{\theta} = \sum_{k=1}^{n} x_n \qquad \hat{\theta} = \frac{1}{n}\sum_{k=1}^{n} \theta_k$$

$$\Rightarrow \boxed{\hat{\theta} = \frac{1}{n}\sum_{k=1}^{n} x_n}$$

Hence Proved !

Ques:-2 for the given normal distribution we know that:

$$P(D|\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \qquad \theta = (\mu, 1)$$

$\therefore$ we can write above eq^n as

$$P(D|\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \rightarrow (1)$$

taking log both the sides ~~and~~ equating ~~it with zero to calculate max likelihood estimate~~

$$\log P(D|\theta) = \sum_{i=1}^{n}\left(\log\frac{1}{\sqrt{2\pi}} + \log e^{-\frac{1}{2}(x_i-\mu)^2}\right) \rightarrow (2)$$

differentiating eq^n (2) and equating it to zero to calculate max. likelihood estimate

$$\frac{\partial}{\partial\mu}\left(\sum_{i=1}^{n} \log\frac{1}{\sqrt{2\pi}}\right) + \frac{\partial}{\partial\mu}\sum_{i=1}^{n}\left(-\frac{1}{2}(x_i-\mu)^2\right) = 0$$

$$\Rightarrow \frac{1}{2} * 2\sum_{i=1}^{n}(x_i-\mu) = 0$$

$$\Rightarrow \frac{1}{n} \sum_{i=1}^{n} x_i = \mu_{mL} \quad \text{--- (3)}$$ this μ value is μ_{mL}

Now if it is given that $n = \infty$ ie data sample is very large.

We know that for class c_1 mean$_{c_1} = 0$
for Class c_2 mean$_{c_2} = 1$ // ground truth given

further c_1 contains x points
c_2 contains $n-x$ points

$\therefore$ Global mean $= \dfrac{\text{mean}_{c_1} * \text{points in } c_1 + \text{mean}_{c_2} * \text{pts in } c_2}{n}$

Global mean $= \dfrac{0 * x + (n-x) * 1}{n} = 1 - \dfrac{x}{n}$

now since $n = \infty$ $\therefore$ global mean $= 1$

Put $n = \infty$ in eq (3)

$$\frac{1}{n} \left(\sum_{i=1}^{\infty} x_i \right) = \mu_{mL}$$

$\sum_{i=1}^{\infty} x_i \xrightarrow[\text{be written as}]{\text{can}}$ $n *$ global_mean
ie $n * 1 = n$

$$\Rightarrow \frac{1}{n} * n = \mu_{mL}$$

$$\Rightarrow \boxed{\mu_{mL} = 1}$$

(b) for Gaussian distribution we know that eq^n of decision boundary is

$$x = \frac{\mu_1 + \mu_2}{2}$$

$\mu_1 = 0$
$\mu_2 = 1$ (calculated in part(a))

$$\therefore \quad x = \frac{1+0}{2} = \frac{1}{2} \Rightarrow \text{eq}^n \ \boxed{x = \frac{1}{2}}$$

(c) $P(x|\omega_1) \sim N(0,1)$
$P(x|\omega_2) \sim N(1, 10^6)$

calculating eq^n of decision boundary

$$P(x|\omega_1)P(\omega_1) = P(x|\omega_2)P(\omega_2)$$

$$\frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} * \cancel{\tfrac{1}{2}} = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}} * \cancel{\tfrac{1}{2}} \qquad \because P(\omega_1) = P(\omega_2) = \tfrac{1}{2}$$

$$\Rightarrow \frac{1}{10^3} e^{-\frac{(x-1)^2}{2\times 10^6}} = \frac{1}{1} e^{-\frac{(x-0)^2}{2}}$$

$\Rightarrow$ taking log both the sides

$$\log\frac{1}{10^3} - \frac{(x-1)^2}{2\times 10^6} = -\frac{x^2}{2}$$

$$-3\log 10 = -\frac{x^2}{2} + \frac{(x-1)^2}{2\times 10^6}$$

$$\Rightarrow x^2(1-10^6) - 2x + 1 + 6*10^6 \log 10 = 0 \quad \longrightarrow (1)$$

Solving above quadratic eq^n

$$x = \frac{+2 \pm \sqrt{4 - 4(1-10^6)(1+6\times 10^6 \log 10)}}{2*(1-10^6)}$$

$x = -3.72, +3.72$

ie

-3.72 $\quad$ $+3.72$

Region B $\quad$ Region A $\quad$ Region C

now we need to check if Region A denotes ω_1 or ω_2

we can check this by taking sample $x=0$ which belongs to region A $[-3.72, 3.72]$ and calculate $P(\omega_1|x=0)$ and $P(\omega_2|x=0)$

$$P(\omega_1|x=0) = \frac{1}{\sqrt{2\pi}*1} e^{-\frac{(0-0)^2}{2*1}} = \frac{1}{\sqrt{2\pi}} * 1$$

$$P(\omega_2|x=0) = \frac{1}{\sqrt{2\pi}*10^3} e^{-\frac{1}{2*10^6}}$$

ie $P(\omega_1|x=0) > P(\omega_2|x=0)$

$\therefore$ Region $[-3.72, 3.72] \in$ class ω_1

we can further check for region B and region C by considering pt $x=-4$ for region B ie $(-\infty, -3.72)$

$$P(\omega_1|x=-4) = \frac{1}{\sqrt{2\pi}*1} e^{-\frac{(-4-0)^2}{2*1}} = \frac{1}{\sqrt{2\pi}} e^{-8}$$

$$P(\omega_2|x=-4) = \frac{1}{\sqrt{2\pi}*10^3} e^{-\frac{(-4-1)^2}{2*10^6}} = \frac{1}{\sqrt{2\pi}*10^3} e^{-\frac{25}{2*10^6}}$$

ie $P(\omega_2|x=-4) > P(\omega_1|x=-4)$

$\therefore$ region B belongs to class ω_2

we can further check for region C by considering pt $x=+4$ which belongs to region C ie $(3.72, \infty)$

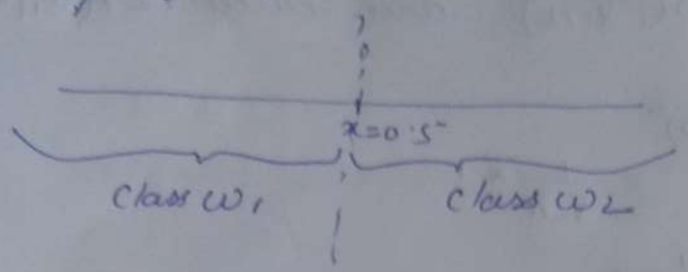
$$P(\omega_1|x=4) = \frac{1}{\sqrt{2\pi}*1} e^{-\frac{(4-0)^2}{2*1}} = \frac{1}{\sqrt{2\pi}} e^{-8}$$

$$P(\omega_2|x=4) = \frac{1}{\sqrt{2\pi}*10^3} e^{-\frac{(4-1)^2}{2*10^3}} = \frac{1}{\sqrt{2\pi}*10^3} e^{-\frac{9}{2*10^3}}$$

ie $P(\omega_2|x=4) > P(\omega_1|x=4)$

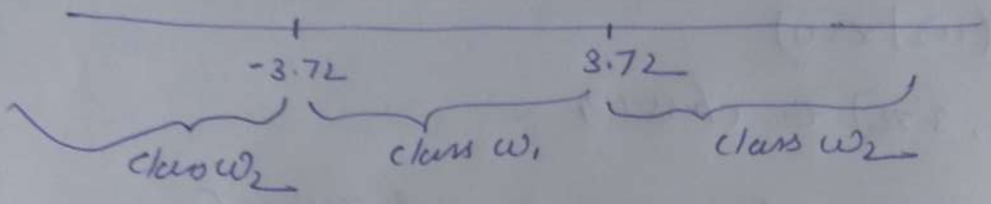
$\therefore$ Region C $\in$ class ω_2

Similarly for part (b) we can show that



→ Poor Model

Part (b)



→ Ground Truth Model

Part (c)

Ques:-3

(a) we already know that if expectation of the estimator equal to true value then the mean estimator is said to be unbiased.

given ~~E(x)~~ x_1 → mean

Please note that I have made an assumption that the given points are iids. [from true distribution]

$E(x_1) = \mu_{ML} = \mu \qquad \because E(x_i)\ \forall i \in [1,N] = E(X) = \mu$

$\therefore$ we can say that this method is unbiased.

(b) This model is nevertheless ~~highly~~ highly undesirable because if x_1 is a sample with very less Probability then the estimated distribution may be very different from the true distribution.

Moreover the estimated [Expected] covariance value will be biased.