

Name - Mahendra Kumar.

Roll no - 1903209

Branch - M&C

Q1

$$\text{given } f_0 \sim N(0, \sigma_0^2)$$

$$f_0(x) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{(x-0)^2}{2\sigma_0^2}} = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{x^2}{2\sigma_0^2}} \quad \text{--- (1)}$$

$$f_1 \sim N(0, \sigma_1^2)$$

$$f_1(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{x^2}{2\sigma_1^2}} \quad \text{--- (2)}$$

$$p_0 = p_1 = 1. \quad \text{there are only two class } y \in \{0, 1\}$$

Bayes classifier

$$h_B(x) = \underset{h}{\operatorname{argmin}} E(L(h(x), y))$$

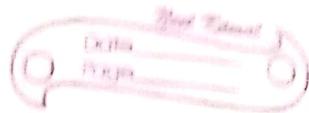
$$E(L(h(x), y)) = E_x(E_{y|x}(L(h(x), y) | x)) \quad \text{--- (1)}$$

$$\therefore E_{y|x}(L(h(x), y) | x) = \sum_{j=0}^1 L(h(x), y=j) q_j(x) \quad \text{--- (2)}$$

$$\text{where } q_j(x) = P(y=j | x=x) = \frac{p_j f_j}{p_0 f_0 + p_1 f_1}$$

Now we want some $h(x) \in \{0, 1\}$ ~~= A~~

$$\text{we want } h_B(x) = \underset{K \in A}{\operatorname{argmin}} \sum_{j=0}^1 L(k, j) q_j(x)$$



by (1)

$$E(L(x), Y) = \int_{\mathbb{R}} E_{Y|x} (L(h(x), Y)|x) \cdot f_x(x)$$

$$\text{so } \arg \min E(L(x), Y) = \arg \min E_{Y|x} (L(x), Y)|x) \\ (\because f_x \geq 0)$$

From (2)

= our bayes classifier is

$$h_B(x) = \underset{k \in A = \{0, 1\}}{\arg \min} \sum_{j=0}^1 L(k, j) q_{kj}(x) \cdot$$

$$= \underset{k \in A}{\arg \min} \begin{cases} L(0, 0) q_0(x) + L(0, 1) q_1(x) & \text{if } k=0 \\ L(1, 0) q_0(x) + L(1, 1) q_1(x) & \text{if } k=1 \end{cases}$$

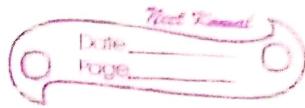
so choose $k = 0 \iff$

$$h_B(x) = \underset{k \in A}{\arg \min} \begin{cases} q_0(x) & \text{if } k=0 \\ q_1(x) & \text{if } k=1 \end{cases} \quad (\because L(i, j) = \begin{cases} 0 & i=j \\ 1 & \text{otherwise} \end{cases})$$

$$\therefore h_B(x) = \begin{cases} 0 & \text{if } q_0(x) \leq q_1(x) \\ 1 & \text{if } q_1(x) > q_0(x) \end{cases}$$

Am. when q_j are called posterior

$$h_B(x)=0 \Rightarrow q_0(x) \geq q_1(x) \Rightarrow \frac{b_0 s_0}{b_0 s_0 + b_1} \geq \frac{b_1 s_1}{b_0 s_0 + b_1 s_1}$$



- Given $b_0 = b_1 = b_2$

- $f_0 \geq f_1$

$$= e^{\frac{1}{\sqrt{2\sigma_1^2}} e^{-\frac{x^2}{2\sigma_1^2}}} \geq e^{\frac{1}{\sqrt{2\sigma_2^2}} e^{-\frac{x^2}{2\sigma_2^2}}}$$

- take log. on both side given

$$-\frac{x^2}{2\sigma_1^2} - \log \sigma_1 \geq -\frac{x^2}{2\sigma_2^2} - \log \sigma_2$$

$$\boxed{x^2 \left(\frac{1}{2\sigma_2^2} - \frac{1}{2\sigma_1^2} \right) \geq \log \left(\frac{\sigma_1}{\sigma_2} \right)} \quad \text{Ans}$$

$$\underline{x^2 \left(\frac{1}{2\sigma_2^2} - \frac{1}{2\sigma_1^2} \right)}$$





Q-2

$$\text{Let } D = \sum_{i=1}^N x_i$$

$$\hat{\mu} = \frac{\sum_{i=1}^N x_i}{N}$$

then

$$E(\hat{\mu}) = E\left(\frac{\sum_{i=1}^N x_i}{N}\right) = \frac{1}{N} \sum_{i=1}^N E(x_i)$$

$$= \frac{1}{N} \left(\underbrace{\mu + \mu + \dots + \mu}_{\text{N times}} \right)$$

$$= \frac{N}{N} \mu = \mu \rightarrow \text{true mean / population mean}$$

$\Rightarrow \hat{\mu}$ is unbiased estimator

$$@ \quad \hat{\sigma}_1^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2$$

$$E(\hat{\sigma}_1^2) = \frac{1}{n} \sum_{i=1}^n E(x_i - \hat{\mu})^2 = \frac{1}{n} \sum_{i=1}^n E(x_i^2 - 2x_i \hat{\mu} + \hat{\mu}^2)$$

$$= \frac{1}{n} \left(\sum_{i=1}^n (E(x_i^2) + E(\hat{\mu}^2)) - 2 \cancel{E(x_i \hat{\mu})} \right)$$

~~($E(x_i \hat{\mu})$)~~
~~($E(\hat{\mu}) = \mu$)~~
~~($E(\mu) = \mu$)~~
~~(μ is constant & given data)~~

$$\Leftrightarrow E(\hat{\mu}^2) = E\left(\left(\frac{\sum x_i}{n}\right) \cdot \left(\frac{\sum x_i}{n}\right)\right)$$

$$= E\left(\sum_{i,j=1}^n \frac{x_i x_j}{n^2}\right) \stackrel{!}{=} E\left(\sum_{i=1}^n \cancel{x_i^2} + \sum_{i \neq j}^n x_i x_j\right)$$

$$n^2 E(\hat{\mu}^2) = \sum_{i=1}^n E(x_i^2) + \sum_{\substack{i,j=1 \\ i \neq j}}^n E(x_i x_j)$$

$\because x_i$'s are iid $\Rightarrow E(x_i x_j)$

$$E(x_i^2) = \sigma^2 - \mu^2 \quad \forall i \quad = E(x_i) E(x_j) \quad \text{if } i \neq j$$

$$\begin{aligned} n^2 E(\hat{\mu}^2) &= \sum_{i=1}^n E(x_i^2) + \sum_{\substack{i,j=1 \\ i \neq j}}^n E(x_i) E(x_j) \\ &= \cancel{\sum_{i=1}^n x_i^2} + \cancel{\sum_{i \neq j} x_i x_j} \quad \sum_{i=1}^n \sum_{j=1}^n \cancel{x_i} \cancel{x_j} E(x_i)^2 \\ \text{but we know. } \& E(x_i^2) = \text{Var}(x_i) + E(x_i)^2 \\ &= \frac{\sigma^2}{n} + \frac{\mu^2}{n} - (1) \\ &\quad \text{Constant} \cancel{+ \frac{1}{n}} \end{aligned}$$

$$n^2 E(\hat{\mu})^2 = n \cancel{(\sigma^2 - \mu^2)} +$$

$$n^2 E(\hat{\mu}) = n \cdot (\sigma^2 - \mu^2) + (n-1)^2 \mu^2$$

$$E(\hat{\mu}^2) = \frac{1}{n} \left(\sum_{i=1}^n (E(x_i^2) + E(\hat{\mu})^2 - 2E(\hat{\mu} x_i)) \right)$$

$$\text{but } E(\hat{\mu} x_i) = E \left(\frac{x_1 + \dots + x_n}{n} x_i \right)$$

$$= \frac{1}{n} E(x_1 x_i) + \dots + E(x_n x_i)$$

$$= \cancel{\frac{1}{n} n E(x_i^2)} + \frac{1}{n} \left(\sigma^2 + (n-1) \frac{(\sigma^2 - \mu^2)}{n} \right)$$

$$= \cancel{\frac{1}{n} n \sigma^2} + \cancel{\frac{1}{n} n \mu^2}$$

$$E(\hat{\mu}^2) = \frac{1}{n} \left(n \cdot (\sigma^2 - \mu^2) + (\sigma^2 - \mu^2) - 2n(\sigma^2 - \mu^2) \right)$$

$$\Rightarrow E(\hat{\sigma}_i^2) = \frac{1}{n} \left(\sum_{i=1}^n E(x_i^2) + \sum_{i=1}^n E(\hat{u}_i^2) - \sum_{i=1}^n E(\hat{u}_i x_i) \right)$$

$$= \frac{1}{n} \left(n(\sigma^2 - \mu^2) + n \underbrace{\left(\frac{n}{n} (\sigma^2 - \mu^2) + (n-1)^2 \mu^2 \right)}_{\left(\frac{n}{n^2} \right)} - 2 \cancel{\left(\frac{1}{n} \left(\cancel{\sigma^2} \mu^2 + (n-1)(\sigma^2 - \mu^2) \right) \right)} \right)$$

$$= \frac{1}{n} \left\{ n(\sigma^2 - \mu^2) + \underbrace{1}_{\frac{1}{n}} (\sigma^2 - \mu^2) + \frac{1}{n} \left(\frac{(n-1)}{n} \right)^2 \mu^2 - 2 (\sigma^2 \mu^2 + (n-1)(\sigma^2 - \mu^2)) \right\}$$

$$= \frac{1}{n} \left(n(\sigma^2 - \mu^2) + (\sigma^2 - \mu^2) + \cancel{\frac{(n-1)}{n} \mu^2} (n-2 + \frac{1}{n}) \mu^2 \right) \xrightarrow{=} 2 \sigma^2 + -2(n-1)(\sigma^2 - \mu^2) - 2\mu^2$$

$$= \frac{1}{n} \cancel{\left((n+1)\sigma^2 \right)} - \cancel{2\mu^2} (2+n-1) - 2(n-1)(\sigma^2 - \mu^2)$$

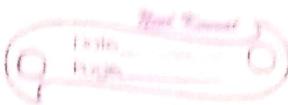
$$= \frac{1}{n} \left((n+1)\sigma^2 - (n+1)\mu^2 + \frac{(n-1)}{n} \mu^2 \right) - 2 ((\sigma^2 - \mu^2) + (n-1)\sigma^2 - (n-1)\mu^2)$$

$$= \frac{1}{n} \left((n+1)(\sigma^2 - \mu^2) - (n+1)\mu^2 + \frac{(n-1)\mu^2}{n} \right)$$

$$- 2 \cancel{(n-1)}$$

"Sorry but I guess I mis calculate something"

OK



on simplification further I go will see.

$$E(\hat{\sigma}_1^2) = \frac{1}{n} \left(\left(\frac{n-1}{n} \right) \sigma^2 \right).$$

(b)

now if we replace $n \rightarrow n-1$

$$\text{we see } E(\hat{\sigma}_2^2) = \frac{1}{n-1} (n-1) \hat{\sigma}^2 = \sigma^2$$

so unbiased estimator.

③

In question ① I did derive a formula
a for 1 & class problem yes, I saw

following along similar line we get

$$h_0(x) = \operatorname{argmin} E(L(h(x), Y))$$

$$= \operatorname{argmin}_{j \in A} E_{Y|X}(L(h(x), Y)|x)$$

$A = \{C_1, C_2, \dots, C_K\} \rightarrow$ all possible classes

$$= \operatorname{argmin}$$

$$E_{Y|X}(L(h(x), Y)|x) = \sum_{j=0}^K L(h(x), C_j) q_j(x)$$

when $q_j(x) = \frac{\text{posterior probability}}{P(Y=j|x=x)}$ → posterior probability

$$= P(Y=j|x=x)$$

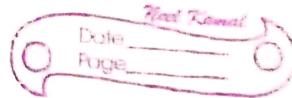
$\Rightarrow h_0(x)$ choose class i s.t. for $x=x$

$$\sum_{l=1}^K E_{Y|X=x}(L(C_l, C_i) | x) \leq \sum_{j=0}^K L(C_m, C_j) q_j(x)$$

$$\sum_{l=1, l \neq i}^K L(C_l, C_i) q_l(x) \leq \sum_{j=0}^K L(C_m, C_j) q_j(x) + L(C_m)$$

$$\Rightarrow \downarrow \quad \because L(C_l, C_i) = 0 \text{ if } l \neq i$$

$$\sum_{l=1, l \neq i}^K L(C_l, C_i) q_l(x) \leq \sum_{j=1}^K L(C_m, C_j) q_j(x)$$



\Rightarrow ~~(E.S)~~

$$1 - q_i(x) \leq 1 - q_m(x)$$

$$\forall i \in \{1, \dots, m\}$$
$$q_i(x) \geq q_m(x)$$

so Bayes classifier with 0-1 loss choose.

label associated with maximum posterior



Q.

X Y

$P(X=x, Y=y)$

1	1	0.3
1	2	0.1
3	1	0.3
2	1	0.15
3	2	0.15

Ans

(a) ~~$P_{X,Y}(x,y) = \sum_{y \in Y} P(X=x, Y=y)$~~ $a=1$

marginal prob. of X will be

X	$P_X(x) = \sum_{y \in Y} P(X=x, Y=y)$
1	$0.1 + 0.3 = 0.4$
2	0.15
3	0.45

marginal .fy

Y	$P_Y(y) = \sum_{x \in X} P(X=x, Y=y)$
1	$0.3 + 0.3 + 0.15 = 0.75$
2	0.25

(6) Conditional distributions.

$$P_{X|Y=y}(x|y) = \frac{P(X=x, Y=y)}{P(Y=y)}$$

$$= \frac{P(X=x \cap Y=y)}{T}$$

~~We can choose $P(Y=y)$~~

Using this formula
we get

X	Y
1	1
1	2
3	1
2	1
3	2

$$P(X|Y) \quad \cancel{P(X=x \cap Y=y)} = P(X=x|Y=y)$$

$$0.3/0.75 = 0.4$$

$$0.1/0.2 = 0.4$$

$$0.3/0.75 = 0.4$$

$$0.15/0.75 = 0.2$$

$$0.15/0.75 = 0.2$$

sim, (only)

$$P_{Y|X=x}(y|x) = \frac{P(X=x, Y=y)}{P(X=x)}$$

X	Y
1	1
1	2
3	1
2	1
3	2

$$P(Y|X) = P(X=x|Y=y)$$

$$0.3/0.4 = 0.75$$

$$0.1/0.4 = 0.25$$

$$0.3/0.45 = 0.666$$

$$0.15/0.15 = 1$$

$$0.15/0.45 = 0.33$$

③ In Regression:

$$h(x) = \hat{w}^T X - w^T X.$$

Under LMS error

$$w^* = (A^T A)^{-1} A^T Y$$

$$Y = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \end{bmatrix} \quad A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 3 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$$

$$A^T A = \left[\begin{array}{ccccc} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 3 & 2 & 3 \end{array} \right] \left[\begin{array}{cc} 1 & 1 \\ 1 & 1 \\ 1 & 3 \\ 1 & 2 \\ 1 & 3 \end{array} \right]$$

$$= \begin{bmatrix} 5 & 10 \\ 10 & 24 \end{bmatrix}$$

$$(A^T A)^{-1} = \frac{1}{20} \begin{pmatrix} 24 & -10 \\ -10 & 5 \end{pmatrix} \quad (\text{Use online calculator})$$

$$(A^T Y) = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 3 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 6 \\ 14 \end{bmatrix}$$

$$\Rightarrow \mathbf{w}^* = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}$$

$$= \begin{bmatrix} 8 & 10 \\ \dots & \dots \\ 20 & 20 \end{bmatrix} \frac{1}{20} \begin{bmatrix} 24 & -10 \\ -10 & 5 \end{bmatrix} \begin{bmatrix} 6 \\ 14 \end{bmatrix}$$

$$\mathbf{w}^* = \frac{1}{20} \begin{pmatrix} 4 \\ 10 \end{pmatrix} = \frac{1}{10} \begin{pmatrix} 2 \\ 5 \end{pmatrix} = \begin{pmatrix} 1/5 \\ 1/2 \end{pmatrix}.$$

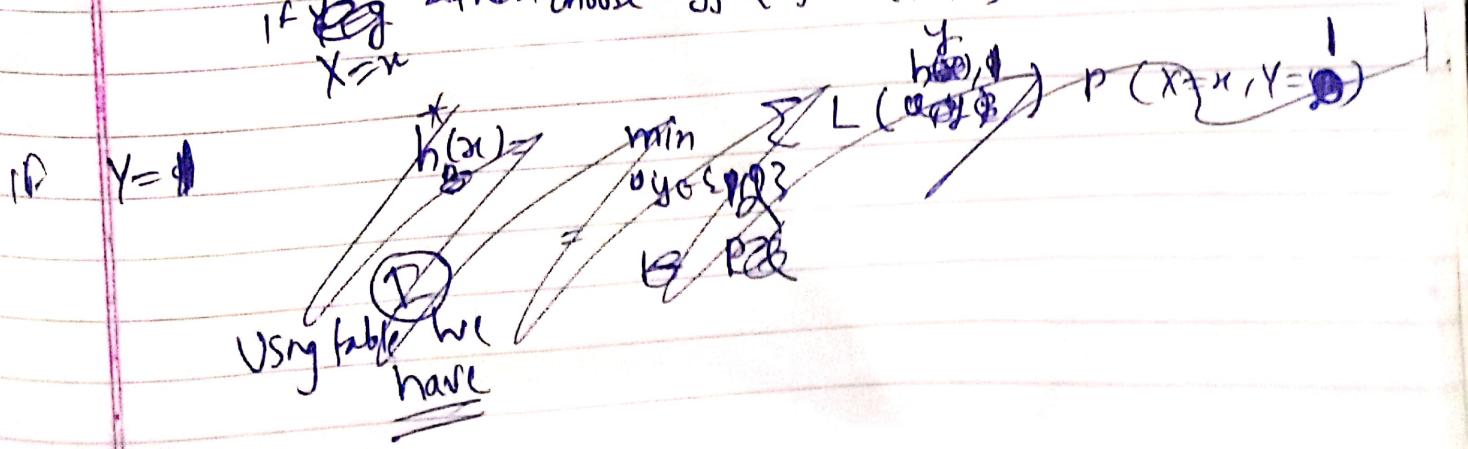
(d) for classification problem,

$$\text{hypo } E(L(h(x), Y)) = \sum_{(x,y) \in \mathcal{D} \cup \mathcal{L}} L(h(x), y) P_{X,Y}(x, y)$$

$$= \sum_{y \in Y} \sum_{x \in X} L(h(x), y) P(x=x, y=y)$$

$$= \sum_{\substack{y \in Y \\ y \in \{1, 2\}}} \left(\sum_{x \in \{1, 2, 3\}} L(h(x), y) P(x=x, y=y) \right)$$

If we choose $h(x) = y_i \min \sum L(y_j, y) p(x=x, y=y)$
 If $x=1$ then choose $y_i \in \{0, 1\} x \in \{1, 2, 3\}$



for

$x=x$

$$h_X^*(x) = \min_{h \in \mathcal{H}} \sum_y L(h(x), y) P(X=x, Y=y)$$

$$= \min_h L(h(x), 1) P(X=x, Y=1) + L(h(x), 2) P(X=x, Y=2)$$

$$= \begin{cases} 1 & \text{if } P(X=x, Y=2) \\ & \leq P(X=x, Y=1) \\ 2 & \text{otherwise.} \end{cases}$$

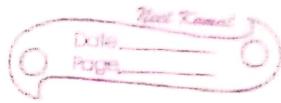
$\Rightarrow h_X^*(x)$ is the required classifier.

For ex so

X	Y	h(x) P(X)	Using above classifier
1	1	h(x) P(X)	h(x) log
1	2	h(x) P(X)	h(x)
2	1	0.3	1 0
2	2	0.1	1 1
3	1	0.3	1 0
3	2	0.15	1 0
		0.15	1 1

$$\text{Error} = 0 + 1 + 0 + 0 + 1$$

$$= \underline{\underline{2}}$$



Q-9

Q-9

Given $A \in \mathbb{R}^{n \times n}$, $C \in \mathbb{R}^{m \times m}$

$$\text{Given } P = \begin{bmatrix} A & U \\ V & C \end{bmatrix}$$

Given C is invertible then

$$R_1 \leftarrow R_1 - UC^{-1}R_2$$

$$\begin{bmatrix} I & -UC^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A & U \\ V & C \end{bmatrix} = \begin{bmatrix} A - UC^{-1}V & 0 \\ 0 & C \end{bmatrix}$$

$$2 \quad \text{To make it } \rightarrow \text{ Reg } C_2 \leftarrow C_2 - VC^{-1}C_1$$

$$\begin{bmatrix} I & -UC^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A & U \\ V & C \end{bmatrix} \begin{bmatrix} I & 0 \\ -VC^{-1} & I \end{bmatrix}$$

$$= \begin{bmatrix} A - UC^{-1}V & 0 \\ 0 & C \end{bmatrix}$$

$$\therefore XYZ = W$$

$$Z^{-1}Y^{-1}X^{-1} = W^{-1}$$

$$Y^{-1} = ZW^{-1}X$$

$$P^{-1} = \begin{bmatrix} A & U \\ V & C \end{bmatrix}^{-1} = \begin{pmatrix} I & 0 \\ -VC^{-1} & I \end{pmatrix} \begin{pmatrix} (A - UC^{-1}V) & 0 \\ 0 & C \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}$$

$$\begin{aligned} P^{-1} &= \begin{pmatrix} I & 0 \\ -vC^T & I \end{pmatrix} \begin{pmatrix} A - vC^T v & (A - vC^T v)(-vC^T) \\ 0 & C \end{pmatrix} \\ &= \begin{pmatrix} A - vC^T v & (A - vC^T v)(-vC^T) \\ -vC^T (A - vC^T v) & -vC^T (A - vC^T v)(-vC^T) + C \end{pmatrix}. \end{aligned}$$

Remaining in next pdf

