Midsem - ML.

Name - Mahendra Kumar.

Roll no - 1903209

Branch - M&C.

Q1 given $f_0 \sim N(0, \sigma_1^2)$

$$f_0(x) = \frac{1}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{(x-0)^2}{2\sigma_1^2}} = \frac{1}{\sqrt{2\pi}\,\sigma_1} e^{\frac{-x^2}{2\sigma_1^2}} \quad -(1)$$

$f_1 \sim N(0, \sigma_2^2)$

$$f_1(x) = \frac{1}{\sqrt{2\pi}\,\sigma_2} e^{\frac{-x^2}{2\sigma_1^2}} \quad -(2)$$

$p_0 = p_1 = \frac{1}{2}$. there are only two class $y \in \{0,1\}$

Bayes classifier

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

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

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

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

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

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

h_1 (1)

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

So $\arg\min E(L(X),Y) = \arg\min E_{Y|X}((L(X),Y)|X)$

$(\because f_X \geq 0)$

from (2)

$\Rightarrow$ our bayes classifier is

$$h_B(x) = k^* = \arg\min_{k\in A=\{0,1\}} \sum_{j=0}^{1} L(k,j)\, q_j(x).$$

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

So ~~choose~~ ~~$k=0$ if.~~

$$h_B(x) = \arg\min_{k\in A} \begin{cases} q_0(x) & \text{if } k=0 \\ q_1(x) & \text{if } k=1 \end{cases}$$

$$\left(\because L(i,j) = \begin{cases} 0 & i=j \\ 1 & \text{otherwise} \end{cases}\right)$$

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

Ans. where q_j are called posterior

if $h_B(x) = 0 \Rightarrow$

$$q_0(x) \geq q_1(x) \Rightarrow \frac{p_0 f_0}{p_0 f_0 + p_1} \geq \frac{p_1 f_1}{p_0 f_0 + p_1 f_1}$$

Date
Page

$\rightarrow$ Given $p_0 = p_1 = p_2$

$\Rightarrow$ $f_0 \geq f_1$

$$\Rightarrow \frac{1}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{x^2}{2\sigma_1^2}} \geq \frac{1}{\sqrt{2\pi}\,\sigma_2} e^{-\frac{x^2}{2\sigma_2^2}}$$

take log on both side give

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

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

$$\cancel{x^2 \geq 2\log\left(\frac{\sigma_1}{\sigma_2}\right)}$$

Q-2 Let $D = \{x_i\}_{i=1}^{N}$

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

then

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

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

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

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

(a)

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

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

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

constant for given data

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

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

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

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

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

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

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} E(x_i)^2$$

but we know $E(x_i^2) = Var(x_i) - E(x_i)^2$

$$= \underline{\sigma^2 - \mu^2} \quad -(1)$$

Constant $\forall i$

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

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

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

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

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

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

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

$$=$$

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

Date ____
Page ____

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

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

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

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

$$\left. - 2\left(\sigma^2 - \mu^2 + (n-1)(\sigma^2 - \mu^2)\right)\right)$$

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

$$\left. - 2\sigma^2 - 2(n-1)(\sigma^2 - \mu^2) - 2\mu^2 \right.$$

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

$$\cancel{+ (n-2)\mu^2 \ldots 2\mu^2}$$

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

$$\left. - 2\left((\sigma^2 - \mu^2) + (n-1)\sigma^2 - (n-1)\mu^2\right)\right)$$

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

$$\cancel{-2(\ldots n}$$

"Sorry sir. I guess I miscalculate something"

OK

On simplification further I will see.

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

(b) Now if we replace $n \to n-1$

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

$= \sigma^2$

So unbiased estimator.

Date
Page

(3) In question (1) I did derive a formula for 1 & 2 class problem. Yes, I saw

following along similar lines we set-

$$h_B(X) = \arg\min E(L(h(X), Y))$$

$$= \underset{j \in A}{\arg\min} \; E_{Y|X}(L(h(X), Y) | X)$$

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

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

where $q_j(x) = \rightarrow$ posterior probability

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

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

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

$$\sum_{l=0}^{K} L(C_l, C_i | x)\, q_i(x) \le \sum_{j=0}^{K} L(C_m, C_j)\, q_j(x) \quad \forall \; C_m$$

$\Rightarrow$ $\because L(C_l, C_i) = 1$ if $l \neq i$

$\Downarrow$

$$\Rightarrow 1 \cdot \sum_{l=1, l \neq i}^{K} 1 \cdot q_i(x) \le \sum_{j=1}^{K} 1 \cdot q_j(x)$$

$$\Rightarrow \quad 1 - q_i(x) \le 1 - q_m(x)$$

$$\Rightarrow \quad \boxed{q_i(x) \ge q_m(x)} \quad \forall\, C_m \in A = \{C_1, \dots, C_K\}$$

So Bayes classifier with 0-1 loss choose label associated with maximum posterior

4.

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

I

(a) ~~marginal P_X $P_X(x) = \sum$ 0.3 + 0.1~~ a = 1

marginal prob. of X will be

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

maggarginal f_Y

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.

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

$$= \frac{\cancel{P(X=x, Y=y)}}{1}$$

~~Now we know~~ P(Y=y)

Using this formula we got

X	Y	P(X|Y) = P(X=x|Y=y)
1	1	$0.3/0.75 = 0.4$
1	2	$0.1/0.25 = 0.4$
3	1	$0.3/0.75 = 0.4$
2	1	$0.15/0.75 = 0.2$
3	2	$0.15/0.75 = 0.2$

Similarly

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

X	Y	P(Y|X) = P(X=x|Y=y)
1	1	$0.3/0.4 = 0.75$
1	2	$0.1/0.4 = 0.25$
3	1	$0.3/0.45 = 0.666$
2	1	$0.15/0.15 = 1$
3	2	$0.15/0.45 = 0.33$

(c) lin Regression.

$$h(x) = \cancel{w^T} x \qquad w^{*T} x.$$

under LMS error

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

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

$$A^T A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 3 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 3 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$$

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

$$(A^T A)^{-1} = \frac{1}{\cancel{100}\ 20} \begin{pmatrix} 24 & -10 \\ -10 & 5 \end{pmatrix}$$ (using online calculator)

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

$\Rightarrow w^* = (A^TA)^{-1}(A^Ty)$

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

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

(d) for classification problem,

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

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

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

If we choose if $X=x$ then choose $h(x) = \min_{y_j\in\{0,1\}} \sum_{x\in\{1,2,3\}} L(y_j,y)\, p(X=x, Y=y)$

Using table we have

for $X = x$

$$h^*_X(x) = \min_{h} \sum_{y} L(h(x), y)\, P(X=x, Y=y)$$

$$= \min_{h} L(h(x),1)\, P(X=x,Y=1) + L(h(x),2)\, P(X=x,Y=2)$$

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

→ $h^*_X(x)$ is the required classifier.

~~Perror~~ So

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

~~h(x)~~ P(X — Using above classifier

X	Y	P(X=x, Y=y)	h(X)	loss
1	1	0.3	1	0
1	2	0.1	1	1
3	1	0.3	1	0
2	1	0.15	1	0
3	2	0.15	1	1

$$P_{error} = 0 + 1 + 0 + 0 + 1$$
$$= \underline{\underline{2}}$$

Q-15

Q-9 Given $A \in \mathbb{R}^{n \times n}$, $C \in \mathbb{R}^{m \times m}$

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

Given C is invertible then
$R_1 \leftarrow R_1 - UC^{-1} R_2$

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

& ~~to make~~ $C_1 \leftarrow C_1 - VC^{-1}C_2$

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

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

$\therefore$ $XYZ = W$

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

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

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

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

$$= \begin{pmatrix} A - UC^{-1}V & (A-UC^{-1}V)(-UC^{-1}) \\ -VC^{-1}(A-UC^{-1}V) & -VC^{-1}(A-UC^{-1}V)(-UC^{-1}) + C \end{pmatrix}.$$

Remaining in next pdf