4(A) Imputed region A_i for class i is defined as

$$A_i = \{x : P(y=i/wX) \geq P(y=j/wX) \ \forall j \in Y\}$$

(i) Imputed region A_i denotes region in space X where probability of class i is higher than probability of any other class $j \in Y$. For this region in space X, we predict class 'i' for multiclass problem.

$$P = \frac{\exp(Xw_i)}{\underbrace{\sum_{j=1}^{K} \exp(Xw_j)}_{\text{nomalization constant}}}$$

log

$$\max \Rightarrow \log P = Xw_i$$

(ii) Note that:

ⓐ $y_i = w_i X$ is monotonous in X hence support following properties:

scaling ⓑ if $\underset{K \in Y}{\arg\max}\ y_k(x)$ is $k = i$ than for $\underset{K \in Y}{\arg\max}(y_k(\lambda x))$ is also at $k = i$ for any +ve constant λ

additive ⓒ if $\underset{K \in Y}{\arg\max}\ y_K(x_1)$ is $k = i$ and

$\underset{K \in Y}{\arg\max}\ y_K(x_2)$ is $k = i$ then

$\underset{K \in Y}{\arg\max}\left(y_K(x_1) + y_K(x_2)\right)$ is also $k = i$

Now, lets take two points x_1 & $x_2 \in A_i$

$\therefore$ $y_i(x_1) \geq y_j(x_1) \quad \forall j \in y$

$y_i(x_2) \geq y_j(x_2) \quad \forall j \in y$

Consider a point $x_\lambda = \lambda x_1 + (1-\lambda)x_2$

~~Consider~~: for $y_k(x_\lambda)$

$\underset{i \in y}{\text{argmax}}$ $= y_k(\lambda x_1 + (1-\lambda) x_2)$

$= \underbrace{\lambda}_{>0} \cdot \underbrace{y_k(x_1)}_{\text{argmax}=i} + \underbrace{(1-\lambda)}_{\geq 0} \underbrace{y_k(x_2)}_{\text{argmax}=i}$

$\therefore \underset{k \in y}{\text{argmax}} \; y_k(x_\lambda)$ is at $k = i$

$\therefore x_\lambda$ also belongs to A_i, $\forall \lambda \in [0,1]$

hence, A_i is a convex set

Answer 4B

~~Answer~~.

$$l(y, Z) = \sum_{i=1}^{k} l^{bin}(\psi(y)_i, Z_i)$$

$$\psi(y)_i = (-1, -1, \ldots \underbrace{1}_{\substack{i^{th} \text{ position} \\ =1}}, -1, \ldots -1) \quad k \text{ elements}$$

$$l(i, \psi(i)) = \sum_{i=1}^{k} l^{bin}(\psi(i), \psi(i))$$

all $\psi(i) \cdot \psi(i) = 1$

$$= (k \cdot) \begin{pmatrix} -1 & -1 \\ -1 & -1 \\ \vdots & \vdots \\ -1 & -1 \\ -1 & -1 \end{pmatrix} \times \log(1 + \exp(-1 \cdot 1))$$

$$= k \cdot \log\left(1 + \frac{1}{e}\right) \quad -①$$

$$l(j, \psi(i)) = \sum_{i=1}^{k} l^{bin}(\psi(j), \psi(i))$$

all $\psi(i) \cdot \psi(i) = 1$, -2; 2 are -1 (i, j)

$$= (k-2) \cdot \log\left(1 + \frac{1}{e}\right) + 2 \cdot \log(1 + \exp(1))$$

$$= (k-2) \log\left(1 + \frac{1}{e}\right) + 2 \log(1 + e) \quad -②$$

$= k \cdot \log\left(1+\frac{1}{e}\right) - 2 \cdot \log\left(1+\frac{1}{e}\right) + 2 \log(1+e)$

$= k \cdot \log\left(1+\frac{1}{e}\right) - 2\,(0.313 - 1.313)$

$\approx k \cdot \log\left(1+\frac{1}{e}\right) + 2\,(+1)$

$= k \log\left(1+\frac{1}{e}\right) + 2 \qquad — (2)$

from (1) & (2) we get

$l(i, \psi(i)) \leq l(j, \psi(i))$

$k \log\left(1+\frac{1}{e}\right) \leq k \log\left(1+\frac{1}{e}\right) + 2 \qquad \forall i, j \in Y$

~~equality holds when~~ equality holds for $i=j$

This indicates that loss function will be minimum at i^{th} class is predicted for $\psi(i)$ → which is aligned with the true value.

Ans 4 (C)

(i) $\psi(i) = \underbrace{[\overbrace{1, 1, \ldots 1}^{i-1}, -1, -1, -1 \ldots -1]}_{(k-1)}$

Same as in part (B),

$l(i, \psi(i)) = \sum_{i=1}^{k-1} l^{bin}(\psi(i)(\psi(i))) = (k-1) \cdot \log\left(1+\frac{1}{e}\right)$ [all predicts are 1]

$l(j, \psi(i)) = \sum_{i=1}^{k-1} l^{bin}(\psi(j) \cdot \psi(i)) = \{k-1-(|i-j|)\} \log\left(1+\frac{1}{e}\right)$

$+ |i-j| \log(1+e)$

$$\begin{array}{cccccc} i=3 & 1 & 1 & -1 & -1 & -1 \\ j=2 & 1 & -1 & -1 & -1 & -1 \end{array} \qquad \begin{array}{cccccc} 1 & 1 & 1 & -1 & -1 & i=4 \\ 1 & -1 & -1 & -1 & -1 & j=2 \end{array}$$

$= (k-1)\log\left(1+\frac{1}{e}\right) + |i-j|$

$= l(i, \psi(i)) + \underbrace{|i-j|}_{+ve\ value}$ (absolute value of $i-j$)

$\therefore$ ~~$l(j, \psi(i)) \geq$~~

$\therefore l(i, \psi(i)) \leq l(j, \psi(i)) \quad \forall i, j \in Y$

(ii) to prove $l(i+1, \psi(i)) \leq l(i+2, \psi(i))$

from part 4 c(i)

LHS we get $l(i+1, \psi(i)) = l(i, \psi(i)) + |i-i+1|$

$= l(i, \psi(i)) + 1 \quad -①$

RHS. $l(i+2, \psi(i)) = l(i, \psi(i)) + |i+2-i|$

$= l(i, \psi(i)) + 2 \quad -②$

from ① & ②

$l(i+1, \psi(i)) \leq l(i+2, \psi(i)) \quad \forall i \in Y$

① This indicates that loss function is min. when class i is predicted for $\psi(i)$ ~~and~~

② loss function puts more penalty ~~as~~ if '$i+2$' class is predicted for actual class 'i' vs. ~~when~~ penalty when '$i+1$' is predicted. That it ~~understands~~ honors ~~the m~~