Date ____/____/____ N° _____

$$P(y_i = k \mid x_i) = p_k(x_i; \beta), \quad \sum_{k=0}^{K-1} p_k(x_i; \beta) = 1.$$

$$p_k(x_i; \beta) = \frac{e^{x_i^T \beta_k}}{\sum_{k=0}^{K-1} e^{x_i^T \beta_l}}$$

$$f(\beta) = \left[-\sum_{i=1}^{n} \left\{ \sum_{k=0}^{K-1} 1(y_i = k) \log p_k(x_i; \beta) \right\} + \frac{\pi}{2} \sum_{k=0}^{K-1} \sum_{j=1}^{p} \beta_{k,j}^2 \right]$$

Damped Newton's update with learning rate $\eta > 0$ has the form;

$$\beta_k^{(t+1)} = \beta_k^{(t)} - \eta \left[\nabla^2 f(\beta_k^{(t)}) \right]^{-1} \nabla f(\beta_k^{(t)}) \quad \text{—} (*)$$

We want to find $\nabla f(\beta_k)$ and $\nabla^2 f(\beta_k)$.

- Compute the gradient.

$$\frac{\partial f(\beta)}{\partial \beta_k} = -\sum_{i=1}^{n} 1(y_i = k) \frac{\partial}{\partial \beta_k} \left[\log p_k(x_i; \beta) \right] + \frac{\pi}{2} \times 2 \beta_k$$

$$= -\sum_{i=1}^{n} 1(y_i = k) \frac{\partial}{\partial \beta_k} \left[\log p_k(x_i; \beta) \right] + \pi \cdot \beta_k \quad \text{—} (1)$$

Consider $\frac{\partial}{\partial \beta_k} \left[\log p_k(x_i; \beta) \right]$

$$\log p_k(x_i; \beta) = \log \frac{e^{x_i^T \beta_k}}{\sum_{l=0}^{K-1} e^{x_i^T \beta_l}} = x_i^T \beta_k - \log \sum_{l=0}^{K-1} e^{x_i^T \beta_l}$$

$$\frac{\partial}{\partial \beta_k} \left(x_i^T \beta_k - \log \sum_{l=0}^{K-1} e^{x_i^T \beta_k} \right)$$

$$= x_i - \frac{1}{\sum_{l=0}^{K-1} e^{x_i^T \beta_l}} \cdot \frac{\partial}{\partial \beta_k} \left(\sum_{l=0}^{K-1} e^{x_i^T \beta_l} \right) \quad \text{—} (2)$$

Date ___/___/___ No ____

Consider $\frac{\partial}{\partial \beta_k} \sum_{l=0}^{k-1} e^{x_i^T \beta_l}$

If $l = k \Rightarrow \frac{\partial}{\partial \beta_k} \sum_{l=0}^{k-1} e^{x_i^T \beta_l} = e^{x_i^T \beta_k} \cdot x_i$

If $l \neq k \Rightarrow \frac{\partial}{\partial \beta_k} \sum_{l=0}^{k-1} e^{x_i^T \beta_l} = 0$

∴ From ②,

$$\frac{\partial}{\partial \beta_k} \log p_k(x_i; \beta) = x_i - \frac{e^{x_i^T \beta_k}}{\sum_{l=0}^{k-1} e^{x_i^T \beta_l}} \cdot x_i$$

$$= x_i - p_k(x_i; \beta) \cdot x_i \quad ; p_k(x_i; \beta) = \frac{e^{x_i^T \beta_k}}{\sum_{l=0}^{k-1} e^{x_i^T \beta_l}}$$

$$= x_i \left[1 - p_k(x_i; \beta)\right]$$

∴ From ①,

$$\frac{\partial f(\beta)}{\partial \beta_k} = -\sum_{i=1}^{n} 1(y_i = k) \cdot x_i \left(1 - p_k(x_i; \beta)\right) + \pi \beta_k$$

$$= -\sum_{i=1}^{n} x_i \left(1(y_i = k) - p_k(x_i; \beta)\right) + \pi \beta_k$$

$$= -X^T \left(1(Y = k) - p_k\right) + \pi \beta_k$$

$$= X^T \left(p_k - 1(Y = k)\right) + \pi \beta_k \quad \text{— ③}$$

$$\frac{\partial^2 f(\beta)}{\partial \beta_k^2} = \frac{\partial}{\partial \beta_k} \left[\sum_{i=1}^{n} x_i \left(p_k(x_i; \beta) - 1(y_i = k)\right)\right] + \pi \quad \text{— ④}$$

Consider $\frac{\partial}{\partial \beta_k} \left[\sum_{i=1}^{n} x_i \, p_k(x_i; \beta)\right]$

$$= \sum_{i=1}^{n} x_i \cdot \frac{\partial}{\partial \beta_k} \left[p_k(x_i; \beta)\right] \quad \text{— ⑤}$$

RATNA

Date ___/___/___ NP ____

⇒ Consider $\frac{\partial}{\partial \beta_K} p_K(x_i;\beta)$

$$= \frac{\partial}{\partial \beta_K} \frac{e^{x_i^T \beta_K}}{\sum_{l=0}^{K-1} e^{x_i^T \beta_l}}$$

$$\frac{\partial}{\partial \beta_K} \frac{e^{x_i^T \beta_K}}{\sum_{l=0}^{K-1} e^{x_i^T \beta_l}} = \frac{\sum_{l=0}^{K-1} e^{x_i^T \beta_l} e^{x_i^T \beta_K} x_i - e^{x_i^T \beta_K} \frac{\partial}{\partial \beta_K} \sum_{l=0}^{K-1} e^{x_i^T \beta_l}}{\left(\sum_{l=0}^{K-1} e^{x_i^T \beta_l}\right)^2}$$

$$= \frac{e^{x_i^T \beta_K}}{\sum_{l=0}^{K-1} e^{x_i^T \beta_l}} x_i - \frac{e^{x_i^T \beta_K} \cdot \frac{\partial}{\partial \beta_K} \sum_{l=0}^{K-1} e^{x_i^T \beta_K}}{\left(\sum_{l=0}^{K-1} e^{x_i^T \beta_l}\right)^2}$$

If $l = K \Rightarrow \frac{e^{x_i^T \beta_K}}{\sum_{l=0}^{K-1} e^{x_i^T \beta_l}} x_i - \frac{e^{x_i^T \beta_K} \cdot e^{x_i^T \beta_K} \cdot x_i}{\left(\sum_{l=0}^{K-1} e^{x_i^T \beta_l}\right)^2}$

$$= p_K(x_i;\beta) \cdot x_i - [p_K(x_i;\beta)]^2 x_i$$

$$= p_K(x_i;\beta)\, x_i\, (1 - p_K(x_i;\beta))$$

If $l \neq K \Rightarrow \frac{e^{x_i^T \beta_K}}{\sum_{l=0}^{K-1} e^{x_i^T \beta_l}} x_i - \frac{e^{x_i^T \beta_K}}{\sum_{l=0}^{K-1} e^{x_i^T \beta_l}} \cdot \sum_{l=0}^{K-1} \frac{e^{x_i^T \beta_l}}{\sum_{l=0}^{K-1} e^{x_i^T \beta_l}} x_i$

$$= p_K(x_i;\beta)\, x_i - p_K(x_i;\beta) \cdot \sum_{l=0}^{K-1} p_l(x_i;\beta) \cdot x_i$$

$$= p_K(x_i;\beta)\, x_i - p_K(x_i;\beta) \cdot x_i \quad ; \sum_{l=0}^{K-1} p_l(x_i;\beta) = 1$$

$$= 0$$

∴ From (5)

$$\frac{\partial}{\partial \beta_K} \left[\sum_{i=1}^{n} x_i\, p_K(x_i;\beta) \right] = \sum_{i=1}^{n} x_i \cdot p_K(x_i;\beta)(1 - p_K(x_i;\beta)) x_i$$

RATHNA

Date ____/____/____ No ____

$$(W_K)_{ii} = p_K(x_i;\beta)\,(1 - p_{Ki}(x_i;\beta))$$

From ④ ;

$$\nabla^2 f = \sum_{i=1}^{n} x_i\, p_K(x_i;\beta)\,(1 - p_K(x_i;\beta))\, x_i + n$$

$$= X^T \left(\sum_{i=1}^{n} W_{Kii}\, x_i\, x_i^T \right) + n I$$

$$= X^T W_K X + n I. \quad — ⑤$$

From ⑧ ;

$$\beta_k^{(t+1)} = \beta_k^{(t)} - \eta \left(X^T W_K X + n I\right)^{-1} \left[X^T \{p_K - 1(Y=k)\} + n \beta_k^{(t)}\right]$$

; $k = 0, \ldots, K-1$

RATHNA