First calculate the gradient :

$$f(\beta) = \left[-\sum_{i=1}^{n} \left\{ \sum_{K=0}^{K-1} 1(y_i = K) \log P_K(x_i) \right\} + \frac{\lambda}{2} \sum_{K=0}^{K-1} \sum_{j=1}^{p} \beta_{kj}^2 \right]$$

$$\nabla_K f(\beta_K) = \frac{\partial f(\beta)}{\partial \beta_K} = \left[-\sum_{i=1}^{n} 1(y_i = k) \frac{\partial}{\partial \beta_K} \log P_K(x_i) + \left(\frac{\lambda}{2}\right) 2\beta_K \right] \quad -(1)$$

$$\frac{\partial}{\partial \beta_K} \log P_K(x_i) = \frac{\partial}{\partial \beta_K} \left[x_i \beta_K - \log \left(\sum_{l=0}^{K-1} e^{x_i \beta_l} \right) \right]$$

$$= \left[x_i - \frac{1}{\sum_{l=0}^{K-1} e^{x_i \beta_l}} \left[e^{x_i \beta_K} \right] x_i \right]$$

$$= \left[x_i - P_K(x_i) x_i \right] = \left[1 - P_K(x_i) \right] x_i \quad -(2)$$

Substitute result of (2) in (1)

$$\nabla_K f(\beta_K) = \left[-\sum_{i=1}^{n} 1(y_i = K) \left[1 - P_K(x_i) \right] x_i + \lambda \beta_K \right] \quad -(3)$$

$$= \left[\sum_{i=1}^{n} \left\{ -1(y_i = k) + 1(y_i = k) P_K(x_i) \right\} x_i + \lambda \beta_K \right]$$

$$\nabla_K f(\beta_K) = \left[X^T \left\{ P_K - 1(Y = K) \right\} + \lambda \beta_K \right] \quad -(4)$$

Next calculate the Hessian

from ③

$$\nabla_K^2 f(\beta_K) = \left[-\sum_{i=1}^{n} 1(y_i = K) \left[-\frac{\partial}{\partial \beta_K} P_K(x_i) \right] x_i + \lambda \right] \quad -⑤$$

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

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

$$= x_i \, P_K(x_i) \left[1 - P_K(x_i) \right] = x_i W_{K\,ii} \quad -⑥$$

Substitute the result from ⑥ in ⑤

$$\nabla_K^2 f(\beta_K) = \left[\sum_{i=1}^{n} 1(y_i = K) \left[x_i W_{K\,ii} \right] x_i + \lambda \right]$$

$$\nabla_K^2 f(\beta_K) = X^T W_K X + \lambda I \quad -⑦$$

Damped Newton's update :

$$\beta_K^{(t+1)} = \beta_K^{(t)} - \eta \left[\nabla_K^2 f(\beta_K) \right]^{-1} \left[\nabla_K f(\beta_K) \right]$$

substituting results from ④ and ⑦

$$\beta_K^{(t+1)} = \beta_K^{(t)} - \eta \left[X^T W_K X + \lambda I \right]^{-1} \left[X^T \left\{ P_K - 1(Y = K) \right\} + \lambda \beta_K^{(t)} \right]$$

for $K = 0, 1, 2, \dots K-1$