(2b)

## Parameter estimation for ordinal likelihood

→ we again use MLE in this case.

From the paper, we get

$$L(B|D) = \prod_{i=1}^{N} \prod_{j=1}^{k} (\pi_j(x_i))^{y_{ij}}$$

$y_{ij} = 1$; if $y_i = j$

$y_{ij} = 0$; Remaing all cases

$$\underbrace{l(\beta|D)}_{\text{say}} = Log(L(\beta|D)) = \sum_{i=1}^{N}\sum_{j=1}^{k} y_{ij} \log(\pi_j(x_i))$$

To maximize, we have to differentiate w.r.t o $\beta$.

we proceed with Netwon - Raphson as our iterative solution for the update rules.

$$B_{n+1} = B_n - H^{-1}(\nabla l(\beta|D))$$

while

$$H = \frac{\partial}{\partial\beta}\left(\frac{\partial}{\partial\beta} l(\beta|D)\right)$$

$$\frac{\partial}{\partial\beta}(l(\beta|D)) = \sum_{i=1}^{N}\sum_{j=1}^{k} y_{ij} \frac{d}{\partial\beta}\underbrace{(\log(\pi_j(x_i)))}_{[\sigma(\theta_j - \beta^T x_i) - \sigma(\theta_{j-1} - \beta^T x_i)]}$$

Solving this we get

$$\boxed{\frac{\partial}{\partial\beta}(l(\beta|D)) = \sum_{i=1}^{N}\sum_{j=1}^{k} y_{ij} x_i [1 - \sigma(\theta_j - \beta^T x_i) - \sigma(\theta_{j-1} - \beta^T x_i)]}$$

After differentiating again, we get the Hessian Matrix.

$$\Rightarrow \quad H = \sum_{i=1}^{N} \sum_{j=1}^{k} y_{ij} (-x_i) \times \frac{\partial}{\partial \beta} \left[ 1 - \sigma(\theta_j - \beta^T x_i) - \sigma(\theta_{j-1} - \beta^T x) \right]$$

$$\boxed{H = \sum_{i=1}^{N} \sum_{j=1}^{k} y_{ij} (-x_i^2) \left[ \sigma(\theta_j - \beta^T x_i) [1 - \sigma(\theta_j - \beta^T x_i)] + \sigma(\theta_{j-1} - \beta^T x_i) [1 - \sigma(\theta_{j-1} - \beta^T x_i)] \right]}$$

using this and gradient, and the update rules, we can estimate the parameters.