

We have:

$$P(Y_i = y_i | x) = p(x_i)^{y_i} (1 - p(x_i))^{1 - y_i} \quad , \quad p(x_i) = \frac{e^{x_i^T \beta}}{1 + e^{x_i^T \beta}}$$

We have:

$$\begin{aligned} l(\beta) &= \log P(Y_1 = y_1, \dots, Y_N = y_N | x_1, \dots, x_N) \\ &\stackrel{\text{ind}}{=} \log \prod_{i=1}^N P(Y_i = y_i | x_i) \\ &= \log \prod_{i=1}^N p(x_i)^{y_i} (1 - p(x_i))^{1 - y_i} \\ &= \sum_{i=1}^N (y_i \log(p(x_i)) + (1 - y_i) \log(1 - p(x_i))) \\ &= \sum_{i=1}^N \left(y_i \underbrace{\log \frac{p(x_i)}{1 - p(x_i)}}_{\substack{= \frac{x_i^T \beta}{e^{x_i^T \beta}} \\ = x_i^T \beta}} + \underbrace{\log(1 - p(x_i))}_{= \frac{1}{1 + e^{x_i^T \beta}} = -\log(1 + e^{x_i^T \beta})} \right) \\ &= \sum_{i=1}^N (y_i x_i^T \beta - \log(1 + e^{x_i^T \beta})) \end{aligned}$$

Further, we have:

$$\begin{aligned} \frac{\partial l(\beta)}{\partial \beta} &= \sum_{i=1}^N \left(y_i x_i - \frac{1}{1 + e^{x_i^T \beta}} \cdot e^{x_i^T \beta} \cdot x_i \right) \\ &= \sum_{i=1}^N \left(y_i - \underbrace{\frac{e^{x_i^T \beta}}{1 + e^{x_i^T \beta}}}_{= p(x_i)} \right) x_i \\ &= \sum_{i=1}^N (y_i - p(x_i)) x_i \end{aligned}$$

$$\text{Let } y = (y_1, \dots, y_N)^T, \quad X = \begin{pmatrix} x_1^T \\ \vdots \\ x_N^T \end{pmatrix}, \quad \beta = (p(x_1), \dots, p(x_N))^T$$

We have:

$$X^T (y - \beta) = (x_1 \dots x_N) \begin{pmatrix} y_1 - p(x_1) \\ \vdots \\ y_N - p(x_N) \end{pmatrix} = \sum_{i=1}^N x_i (y_i - p(x_i)) = \frac{\partial l(\beta)}{\partial \beta}$$

Moreover, we have:

$$\begin{aligned} \frac{\partial^2 l(\beta)}{\partial \beta \partial \beta^T} &= - \sum_{i=1}^N \frac{e^{x_i^T \beta} x_i (1 + e^{x_i^T \beta}) - e^{x_i^T \beta} \cdot e^{x_i^T \beta} \cdot x_i}{(1 + e^{x_i^T \beta})^2} x_i^T \\ &= - \sum_{i=1}^N \frac{e^{x_i^T \beta} + e^{2x_i^T \beta} - e^{2x_i^T \beta}}{(1 + e^{x_i^T \beta})^2} x_i x_i^T \\ &= - \sum_{i=1}^N \frac{e^{x_i^T \beta}}{(1 + e^{x_i^T \beta})^2} x_i x_i^T \\ &= - \sum_{i=1}^N \underbrace{\frac{e^{x_i^T \beta}}{1 + e^{x_i^T \beta}}}_{p(x_i)} \cdot \underbrace{\frac{1}{1 + e^{x_i^T \beta}}}_{1 - p(x_i)} \cdot x_i x_i^T \\ &= - \sum_{i=1}^N p(x_i) (1 - p(x_i)) x_i x_i^T \end{aligned}$$

$$\text{Let } W = \begin{pmatrix} p(x_1)(1 - p(x_1)) & & 0 \\ & \ddots & \\ 0 & & p(x_N)(1 - p(x_N)) \end{pmatrix}$$

$$\begin{aligned} X^T W X &= (x_1 \dots x_N) \begin{pmatrix} p(x_1)(1 - p(x_1)) & & 0 \\ & \ddots & \\ 0 & & p(x_N)(1 - p(x_N)) \end{pmatrix} \begin{pmatrix} x_1^T \\ \vdots \\ x_N^T \end{pmatrix} \\ &= (x_1 \dots x_N) \cdot \begin{pmatrix} p(x_1)(1 - p(x_1)) x_1^T \\ \vdots \\ p(x_N)(1 - p(x_N)) x_N^T \end{pmatrix} = \sum_{i=1}^N p(x_i) (1 - p(x_i)) x_i x_i^T = - \frac{\partial^2 l(\beta)}{\partial \beta \partial \beta^T} \end{aligned}$$

A Taylor expansion of $\frac{\partial l(\beta)}{\partial \beta}$ around β^* is given by

$$\frac{\partial l(\beta)}{\partial \beta} \approx \frac{\partial l(\beta^*)}{\partial \beta} + \frac{\partial^2 l(\beta^*)}{\partial \beta \partial \beta^T} (\beta - \beta^*)$$

Setting $\frac{\partial l(\beta)}{\partial \beta} = 0$, we obtain:

$$\frac{\partial l(\beta^*)}{\partial \beta} + \frac{\partial^2 l(\beta^*)}{\partial \beta \partial \beta^T} (\beta - \beta^*) \approx 0$$

$$\beta - \beta^* \approx - \left(\frac{\partial^2 l(\beta^*)}{\partial \beta \partial \beta^T} \right)^{-1} \frac{\partial l(\beta^*)}{\partial \beta}$$

$$\beta \approx \beta^* - \left(\frac{\partial^2 l(\beta^*)}{\partial \beta \partial \beta^T} \right)^{-1} \frac{\partial l(\beta^*)}{\partial \beta}$$