

We have:

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

We get:

$$\begin{aligned} l(\beta) &= \log P(Y_1 = y_1, \dots, Y_N = y_N | \underline{x}_1, \dots, \underline{x}_N) \\ &\stackrel{\text{ind.}}{=} \log \prod_{i=1}^N P(Y_i = y_i | \underline{x}_i) \\ &= \log \prod_{i=1}^N p(\underline{x}_i)^{y_i} (1 - p(\underline{x}_i))^{1-y_i} \\ &= \sum_{i=1}^N \log (p(\underline{x}_i)^{y_i} (1 - p(\underline{x}_i))^{1-y_i}) \\ &= \sum_{i=1}^N (\log p(\underline{x}_i)^{y_i} + \log ((1 - p(\underline{x}_i))^{1-y_i})) \\ &= \sum_{i=1}^N (y_i \log(p(\underline{x}_i)) + (1 - y_i) \log(1 - p(\underline{x}_i))) \\ &= \sum_{i=1}^N (y_i (\log(p(\underline{x}_i)) - \log(1 - p(\underline{x}_i))) + \log(1 - p(\underline{x}_i))) \\ &\quad \underbrace{\log\left(\frac{p(\underline{x}_i)}{1 - p(\underline{x}_i)}\right)}_{= e^{\underline{x}_i^T \beta}} \\ &\quad \underbrace{\underline{x}_i^T \beta}_{\underline{x}_i^T \beta} \\ &= \sum_{i=1}^N \left(y_i \underline{x}_i^T \beta + \log \left(1 - \frac{e^{\underline{x}_i^T \beta}}{1 + e^{\underline{x}_i^T \beta}} \right) \right) \\ &\quad = \frac{1}{1 + e^{\underline{x}_i^T \beta}} = -\log(1 + e^{\underline{x}_i^T \beta}) \\ &= \sum_{i=1}^N (y_i \underline{x}_i^T \beta - \log(1 + e^{\underline{x}_i^T \beta})) \end{aligned}$$

Further, we have:

$$\frac{\partial l(\beta)}{\partial \beta} = \sum_{i=1}^N \left(y_i \underline{x}_i - \underbrace{\frac{1}{1 + e^{\underline{x}_i^T \beta}} e^{\underline{x}_i^T \beta}}_{p(\underline{x}_i)} \cdot \underline{x}_i \right) = \sum_{i=1}^N (y_i - p(\underline{x}_i)) \underline{x}_i$$

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

Let
$$\underline{W} = \begin{pmatrix} p(\underline{x}_1)(1 - p(\underline{x}_1)) & & & \\ p(\underline{x}_2)(1 - p(\underline{x}_2)) & & & \\ & & \ddots & \\ & & & p(\underline{x}_N)(1 - p(\underline{x}_N)) \end{pmatrix}$$

$$\begin{aligned} \underline{X}^T \underline{W} \underline{X} &= (\underline{x}_1 \dots \underline{x}_N) \begin{pmatrix} p(\underline{x}_1)(1 - p(\underline{x}_1)) & & & \\ & \ddots & & \\ & & \ddots & \\ & & & p(\underline{x}_N)(1 - p(\underline{x}_N)) \end{pmatrix} \begin{pmatrix} \underline{x}_1^T \\ \vdots \\ \underline{x}_N^T \end{pmatrix} \\ &= (\underline{x}_1 \dots \underline{x}_N) \begin{pmatrix} p(\underline{x}_1)(1 - p(\underline{x}_1)) \underline{x}_1^T \\ \vdots \\ p(\underline{x}_N)(1 - p(\underline{x}_N)) \underline{x}_N^T \end{pmatrix} \\ &= \sum_{i=1}^N p(\underline{x}_i)(1 - p(\underline{x}_i)) \underline{x}_i \underline{x}_i^T \end{aligned}$$

A first order 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

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

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

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

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