5. For the diagonal of $A_E(\vec{w})$, we have $\frac{\partial^2}{\partial w_j \partial w_i} E(\vec{w})$ $i,j = 1,\dots,d$

$$\frac{\partial^2}{\partial w_j \partial w_i} E(\vec{w}) = \frac{\partial^2}{\partial w_j \partial w_i} E_{in}(\vec{w})$$

$$= \frac{\partial^2}{\partial w_j \partial w_i} \frac{1}{N} \sum_{n=1}^{N} \ln(1+e^{-y_n \vec{w}^T \vec{x}_n})$$

$$= \frac{1}{N} \frac{\partial}{\partial w_j} \left(\underline{\frac{\partial}{\partial w_i} \sum_{n=1}^{N} \ln(1+e^{-y_n \vec{w}^T \vec{x}_n})} \right) - *$$

first calculate this part.

Let $f_n = \ln(1+e^{-y_n \vec{w}^T \vec{x}_n})$

$$\frac{\partial}{\partial w_i} \ln(1+e^{-y_n \vec{w}^T \vec{x}_n}) = \frac{\partial f_n}{\partial w_i}$$

Let $g_n = -y_n \vec{w}^T \vec{x}_n$

$\frac{d f_n}{d g_n} \cdot \frac{\partial g_n}{\partial w_i}$

$$= \frac{1}{1+e^{g_n}} (-y_n x_{n,i})$$

where $x_{n,i}$ is the i-th term of $\vec{x}_n$

$$= \frac{1}{1+e^{-y_n \vec{w}^T \vec{x}_n}} (-y_n x_{n,i})$$

Plug in to $*$:

$$= -\frac{1}{N} \frac{\partial}{\partial w_j} \left(\sum_{n=1}^{N} \frac{y_n x_{n,i}}{1+e^{-y_n \vec{w}^T \vec{x}_n}} \right)$$

$$= -\frac{1}{N} \sum_{n=1}^{N} \frac{\partial}{\partial w_j} \left(\frac{y_n x_{n,i}}{1+e^{-y_n \vec{w}^T \vec{x}_n}} \right)$$

$$= -\frac{1}{N} \sum_{n=1}^{N} y_n x_{n,i} \cdot \frac{\partial}{\partial w_j} \left(\frac{1}{1+e^{-y_n \vec{w}^T \vec{x}_n}} \right)$$

$$= -\frac{1}{N} \sum_{n=1}^{N} y_n x_{n,i} \cdot \frac{-y_n x_{n,j} \cdot e^{-y_n \vec{w}^T \vec{x}_n}}{(1+e^{-y_n \vec{w}^T \vec{x}_n})^2}$$

$$= \frac{1}{N} \sum_{n=1}^{N} \frac{y_n^2 x_{n,i} x_{n,j} e^{-y_n \vec{w}^T \vec{x}_n}}{(1+e^{-y_n \vec{w}^T \vec{x}_n})^2}$$

is the i,j-th $(j,i$-th$)$ term of the Hessian $A_E(\vec{w})$

$$A_E(\vec{w}) \in \mathbb{R}^{d \times d}$$
$$\parallel$$
$$\underset{N\times N}{X^T D} \underset{N\times d}{X}$$
$$X = \begin{bmatrix} \vec{x}_1^T \\ \vec{x}_2^T \\ \vdots \\ \vec{x}_N^T \end{bmatrix}$$

If we further decompose the result, we can obtain:

$$\frac{1}{N}\sum_{n=1}^{N} \frac{y_n^2 x_{n,i} x_{n,j} e^{-y_n \vec{w}^T \vec{x}_n}}{(1+e^{-y_n \vec{w}^T \vec{x}_n})^2}$$

$$= \frac{1}{N}\sum_{n=1}^{N} x_{n,i} x_{n,j} \frac{y_n^2 e^{-y_n \vec{w}^T \vec{x}_n}}{(1+e^{-y_n \vec{w}^T \vec{x}_n})^2}$$

$\because$ Let $\vec{x}_n = [x_{n1}\ x_{n2} \cdots x_{nd}]^T$

$$\vec{x}_n \vec{x}_n^T = \begin{bmatrix} x_{n1} \\ \vdots \\ x_{nd} \end{bmatrix} [x_{n1} \cdots x_{nd}] = \begin{bmatrix} x_{n1}^2 & x_{n1}x_{n2} & \cdots & x_{n1}x_{nd} \\ x_{n2}x_{n1} & x_{n2}^2 & & \vdots \\ \vdots & & \ddots & \\ x_{nd}x_{n1} & \cdots & & x_{nd}^2 \end{bmatrix}$$

so that:

$$A_E(\vec{w}_t) = \frac{1}{N}\sum_{n=1}^{N} \boxed{\frac{y_n^2 e^{-y_n \vec{w}_t^T \vec{x}_n}}{(1+e^{-y_n \vec{w}_t^T \vec{x}_n})^2}} \vec{x}_n \vec{x}_n^T$$

$= l_n$

$$= \frac{1}{N}(l_1 \vec{x}_1 \vec{x}_1^T + l_2 \vec{x}_2 \vec{x}_2^T + \cdots + l_N \vec{x}_N \vec{x}_N^T)$$

$$= \vec{x}_1 (\tfrac{1}{N} l_1) \vec{x}_1^T + \vec{x}_2 (\tfrac{1}{N} l_2) \vec{x}_2^T + \cdots + \vec{x}_N (\tfrac{1}{N} l_N) \vec{x}_N^T$$

$$= [\vec{x}_1 \cdots \vec{x}_N] \begin{bmatrix} \frac{1}{N} l_1 & & 0 \\ & \frac{1}{N} l_2 & \\ 0 & & \ddots \\ & & & \frac{1}{N} l_N \end{bmatrix}_{N\times N} \begin{bmatrix} \vec{x}_1^T \\ \vdots \\ \vec{x}_N^T \end{bmatrix}$$

$\because X = \begin{bmatrix} \vec{x}_1^T \\ \vdots \\ \vec{x}_N^T \end{bmatrix}$

$$= X^T D X$$

$\therefore X^T = [\vec{x}_1 \cdots \vec{x}_N]$

where $D = \begin{bmatrix} \frac{1}{N} l_1 & & 0 \\ & \frac{1}{N} l_2 & \\ 0 & & \ddots \\ & & & \frac{1}{N} l_N \end{bmatrix}_{N\times N}$ is a diagonal matrix

$$l_n = \frac{y_n^2 e^{-y_n \vec{w}_t^T \vec{x}_n}}{(1+e^{-y_n \vec{w}_t^T \vec{x}_n})^2} = \frac{y_n^2 e^{-y_n \vec{w}_t^T \vec{x}_n}}{h_t(y_n \vec{x}_n)^2} \quad \square$$

$$\because \theta(s) = \frac{1}{1+e^{-s}}$$

$$h_t(\vec{x}) = \theta(\vec{w}_t^T \vec{x}) = (1+e^{-\vec{w}_t^T \vec{x}})^{-1}$$

$$\rightarrow h_t(y_n \vec{x}_n) = \theta(y_n \vec{w}_t^T \vec{x}_n) = (1+e^{-y_n \vec{w}_t^T \vec{x}_n})$$