Evaluation Criterion

$$\mathcal{E}(\boldsymbol{\beta}) = -\sum_{n=1}^{N} \left\{ y_n \log\left[\sigma(\boldsymbol{\beta}^T \mathbf{x_n})\right] + (1-y_n)\log\left[1-\sigma(\boldsymbol{\beta}^T \mathbf{x_n})\right]\right\}$$

No closed-form solution that minimizes the function.
We use an approximate method, e.g. gradient descent, so we need to compute $\nabla\mathcal{E}(\boldsymbol{\beta})$.

Derivatives of sigmoid function $\sigma(\eta)$

$$\sigma(\eta) = \frac{1}{1+e^{-n}}$$

$$\frac{d\sigma(\eta)}{d\eta} = -\frac{-e^{-n}}{(1+e^{-n})^2} = \frac{e^{-n}}{(1+e^{-n})^2} = \frac{1}{1+e^{-n}}\left(\frac{e^{-n}}{1+e^{-n}}\right) = \frac{1}{1+e^{-n}}\left(1-\frac{1}{1+e^{-n}}\right) = \sigma(\eta)\left[1-\sigma(\eta)\right]$$

$$\frac{d\log\sigma(\eta)}{d\eta} = \frac{1}{\sigma(\eta)} \cdot \frac{d\sigma(\eta)}{d\eta} = 1-\sigma(\eta)$$

Derivation of $\nabla\mathcal{E}(\boldsymbol{\beta}) = \frac{\vartheta\mathcal{E}(\boldsymbol{\beta})}{\vartheta\boldsymbol{\beta}}$

$$\begin{aligned}
\nabla\mathcal{E}(\boldsymbol{\beta}) &= -\sum_{n=1}^{N}\left\{y_n\left[1-\sigma(\boldsymbol{\beta}^T\mathbf{x_n})\right]\mathbf{x_n} - (1-y_n)\left[1-\big(1-\sigma(\boldsymbol{\beta}^T\mathbf{x_n})\big)\right]\mathbf{x_n}\right\} \\
&= -\sum_{n=1}^{N}\left\{y_n\left[1-\sigma(\boldsymbol{\beta}^T\mathbf{x_n})\right]\mathbf{x_n} + (1-y_n)\sigma(\boldsymbol{\beta}^T\mathbf{x_n})\mathbf{x_n}\right\} \\
&= -\sum_{n=1}^{N}\left[y_n - y_n\sigma(\boldsymbol{\beta}^T\mathbf{x_n}) - \sigma(\boldsymbol{\beta}^T\mathbf{x_n}) + y_n\sigma(\boldsymbol{\beta}^T\mathbf{x_n})\right]\mathbf{x_n} \\
&= \sum_{n=1}^{N}\underbrace{\big(\sigma(\boldsymbol{\beta}^T\mathbf{x_n}) - y_n\big)}_{\text{error}}\mathbf{x_n}
\end{aligned}$$

Derivation of $\mathbf{H} = \frac{\vartheta^2\mathcal{E}(\boldsymbol{\beta})}{\vartheta\boldsymbol{\beta}\boldsymbol{\beta}^T}$

$$\begin{aligned}
\mathbf{H} &= \frac{\vartheta^2}{\vartheta\boldsymbol{\beta}\boldsymbol{\beta}^T}\left[\sum_{n=1}^{N}\big(\sigma(\boldsymbol{\beta}^T\mathbf{x_n})\cdot\mathbf{x_n} - y_n\mathbf{x_n}\big)\right] \\
&= \sum_{n=1}^{N}\underbrace{\sigma(\boldsymbol{\beta}^T\mathbf{x_n})}_{\in[0,1]}\cdot\underbrace{\big(1-\sigma(\boldsymbol{\beta}^T\mathbf{x_n})\big)}_{\in[0,1]}\cdot\underbrace{\big(\mathbf{x_n}\cdot\mathbf{x_n}^T\big)}_{\in\mathcal{R}^{D\times D}}
\end{aligned}$$