Evaluation Criterion

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

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

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 $\nabla \mathcal{E}(\mathbf{w}) = \frac{\vartheta \mathcal{E}(\mathbf{w})}{\vartheta \mathbf{w}}$

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

$$= -\sum_{n=1}^{N} \left\{ y_n \left[1 - \sigma(\mathbf{w}^T \mathbf{x_n})\right] \mathbf{x_n} + (1 - y_n)\sigma(\mathbf{w}^T \mathbf{x_n})\mathbf{x_n} \right\}$$

$$= -\sum_{n=1}^{N} \left[y_n - y_n\sigma(\mathbf{w}^T \mathbf{x_n}) - \sigma(\mathbf{w}^T \mathbf{x_n}) + y_n\sigma(\mathbf{w}^T \mathbf{x_n}) \right] \mathbf{x_n}$$

$$= \sum_{n=1}^{N} \underbrace{\left(\sigma(\mathbf{w}^T \mathbf{x_n}) - y_n\right)}_{\text{error}} \mathbf{x_n}$$

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

$$\mathbf{H} = \frac{\vartheta^2}{\vartheta \mathbf{w} \mathbf{w}^T}\left[\sum_{n=1}^{N}\left(\sigma(\mathbf{w}^T \mathbf{x_n}) \cdot \mathbf{x_n} - y_n\mathbf{x_n}\right)\right]$$

$$= \sum_{n=1}^{N} \underbrace{\sigma(\mathbf{w}^T \mathbf{x_n})}_{\in [0,1]} \cdot \underbrace{\left(1 - \sigma(\mathbf{w}^T \mathbf{x_n})\right)}_{\in [0,1]} \cdot \underbrace{\left(\mathbf{x_n} \cdot \mathbf{x_n}^T\right)}_{\in \mathcal{R}^{D \times D}}$$