5.19 For a network having a single output with a logistic sigmoid output-unit activation function and a cross-entropy error function (given by 4.90):

$$E = -\sum_{n=1}^{N} \{t_n \ln y_n + (1 - t_n) \ln(1 - y_n)\}$$

$$\nabla E = -\sum_{n=1}^{N} \left\{ \left(\frac{t_n}{y_n}\right) \nabla y_n + \frac{(1 - t_n)}{(1 - y_n)}(-\nabla y_n) \right\}$$

$$= \sum_{n=1}^{N} \left\{ \frac{(1 - t_n)}{(1 - y_n)} \nabla y_n - \left(\frac{t_n}{y_n}\right) \nabla y_n \right\}$$

For 2 class classification, $y_n = \sigma(a_n)$, and $\nabla y_n = \sigma(a_n)(1 - \sigma(a_n))\nabla a_n = y_n(1 - y_n)\nabla a_n$.

Note: We have to think in terms of a_n because as seen in page 251 of the book, $\mathbf{b}_n = \nabla a_n$.

$$\implies \nabla E = \sum_{n=1}^{N} \left\{ \frac{(1 - t_n)}{(1 - y_n)}(y_n(1 - y_n)\nabla a_n) - \left(\frac{t_n}{y_n}\right)(y_n(1 - y_n)\nabla a_n) \right\}$$

$$= \sum_{n=1}^{N} \{y_n(1 - t_n)\nabla a_n - t_n(1 - y_n)\nabla a_n\}$$

$$= \sum_{n=1}^{N} \{y_n(1 - t_n) - t_n(1 - y_n)\} \nabla a_n$$

$$= \sum_{n=1}^{N} \{y_n - t_n\} \nabla a_n$$

$$\implies \mathbf{H} = \nabla\nabla E = \nabla \left(\sum_{n=1}^{N} \{y_n - t_n\} \nabla a_n \right)$$

$$= \sum_{n=1}^{N} \nabla y_n \nabla a_n + \sum_{n=1}^{N} \{y_n - t_n\} \nabla\nabla a_n$$

Neglecting the second term as per the Levenberg–Marquardt approximation,

$$\mathbf{H} \simeq \sum_{n=1}^{N} y_n(1 - y_n)\nabla a_n \nabla a_n$$

$$\simeq \sum_{n=1}^{N} y_n(1 - y_n)\mathbf{b}_n \mathbf{b}_n^T$$