5.16 In case of multiple outputs, the error function is given by:

$$E = \frac{1}{2}\sum_{n=1}^{N} ||\mathbf{y}_n - \mathbf{t}_n||^2 = \frac{1}{2}\sum_{n=1}^{N}(\mathbf{y}_n - \mathbf{t}_n)^T(\mathbf{y}_n - \mathbf{t}_n)$$

The first derivative is given by:

$$\triangledown E = \sum_{n=1}^{N} \triangledown \mathbf{y}_n^T(\mathbf{y}_n - \mathbf{t}_n)$$

Assuming K outputs, $\mathbf{y}_n$ has K elements, and assuming $\mathbf{w}$ has W elements, $\triangledown \mathbf{y}_n$ will be a matrix of size $K * W$.

Let $\mathbf{B}_n = \triangledown \mathbf{y}_n$.

The second derivative is given by:

$$\begin{aligned}\triangledown\triangledown E &= \left(\sum_{n=1}^{N} \triangledown \mathbf{y}_n^T \triangledown \mathbf{y}_n\right) + \left(\sum_{n=1}^{N}(\mathbf{y}_n - \mathbf{t}_n)^T \triangledown\triangledown \mathbf{y}_n\right) \\ &= \left(\sum_{n=1}^{N} \mathbf{B}_n^T \mathbf{B}_n\right) + \left(\sum_{n=1}^{N}(\mathbf{y}_n - \mathbf{t}_n)^T \triangledown\triangledown \mathbf{y}_n\right)\end{aligned}$$

Neglecting the second term as done in the book, we get:

$$\implies \mathbf{H} \simeq \sum_{n=1}^{N} \mathbf{B}_n^T \mathbf{B}_n$$