5.22 Note: j' and j refer to 2 indices for the same set of nodes, not different indices for the 2 hidden nodes. Same for i, i' and k, k'.

$$a_j = \sum_i w_{ji} x_i$$

$$z_j = h(a_j)$$

$$a_k = \sum_j w_{kj} z_j$$

1. Both weights in the second layer:

Using 5.50:

$$\frac{\partial E_n}{\partial w_{kj}^{(2)}} = \left(\frac{\partial E_n}{\partial a_k}\right)\left(\frac{\partial a_k}{\partial w_{kj}^{(2)}}\right) = \delta_k z_j$$

$$\begin{aligned}
\implies \frac{\partial^2 E_n}{\partial w_{kj}^{(2)} \partial w_{k'j'}^{(2)}} &= z_j \left(\frac{\partial^2 E_n}{\partial w_{k'j'}^{(2)}\ \partial a_k}\right) \\
&= z_j \left(\frac{\partial^2 E_n}{\partial a_{k'} \partial a_k}\ \frac{\partial a_{k'}}{\partial w_{k'j'}^{(2)}}\right) \\
&= z_j \left(\frac{\partial^2 E_n}{\partial a_{k'}\ \partial a_k}\ z_{j'}\right) \\
&= z_j z_{j'} M_{kk'}
\end{aligned}$$

which is the same as 5.93.

2. Both weights in the first layer:

Using eqn. 5.50:

$$\begin{aligned}
\frac{\partial E_n}{\partial w_{ji}^{(1)}} &= \frac{\partial E_n}{\partial a_j}\ \frac{\partial a_j}{\partial w_{ji}^{(1)}} \\
&= \delta_j z_i
\end{aligned}$$

$$= \delta_j x_i$$

because z_i refers to the output of i^{th} node

Using 5.56:

$$\delta_j = h'(a_j) \sum_k w_{kj}^{(2)} \delta_k$$

$$\implies \frac{\partial E_n}{\partial w_{ji}^{(1)}} = x_i \, h'(a_j) \sum_k w_{kj}^{(2)} \delta_k$$

$$\implies \frac{\partial^2 E_n}{\partial w_{j'i'}^{(1)} \, \partial w_{ji}^{(1)}} = x_i \left(\frac{\partial h'(a_j)}{\partial w_{j'i'}^{(1)}} \right) \left(\sum_k w_{kj}^{(2)} \delta_k \right) + x_i \, h'(a_j) \left(\sum_k w_{kj}^{(2)} \left(\frac{\partial \delta_k}{\partial w_{j'i'}^{(1)}} \right) \right)$$

The derivative in the first term can be simplified as:

$$\frac{\partial h'(a_j)}{\partial w_{j'i'}^{(1)}} = \frac{\partial h'(a_j)}{\partial a_j} \frac{\partial a_j}{\partial w_{ji'}^{(1)}} = x_{i'} \, h''(a_j) \text{ if } j = j', 0 \text{ otherwise.}$$

The derivative in the second term can be simplified as:

$$\begin{aligned} \frac{\partial \delta_k}{\partial w_{j'i'}^{(1)}} &= \frac{\partial^2 E_n}{\partial w_{j'i'}^{(1)} \, \partial a_k} \\ &= \frac{\partial}{\partial a_k} \frac{\partial E_n}{\partial w_{j'i'}^{(1)}} \end{aligned}$$

Using 5.50, this becomes:

$$= \frac{\partial}{\partial a_k} \frac{\partial E_n}{\partial a_{j'}} \frac{\partial a_{j'}}{\partial w_{j'i'}^{(1)}}$$

Now applying 5.55, we get:

$$= \frac{\partial}{\partial a_k} \left(\sum_{k'} \frac{\partial E_n}{\partial a_{k'}} \frac{\partial a_{k'}}{\partial a_{j'}} \right) \frac{\partial a_{j'}}{\partial w^{(1)}_{j'i'}}$$

$$= \frac{\partial}{\partial a_k} \left(\sum_{k'} \frac{\partial E_n}{\partial a_{k'}} \left(h'(a_{j'})\, w^{(2)}_{k'j'} \right) \right) x_{i'}$$

$$= x_{i'}\, h'(a_{j'}) \sum_{k'} \left(\frac{\partial^2 E_n}{\partial a_k\, \partial a_{k'}} \right) w^{(2)}_{k'j'}$$

$$= x_{i'}\, h'(a_{j'}) \sum_{k'} M_{kk'}\, w^{(2)}_{k'j'}$$

Substituting the two derivatives back, we get:

$$\implies \frac{\partial^2 E_n}{\partial w^{(1)}_{j'i'}\, \partial w^{(1)}_{ji}} = x_i\, x_{i'}\, I_{jj'} h''(a_j) \left(\sum_k w^{(2)}_{kj}\, \delta_k \right)$$

$$+ x_i\, h'(a_j) \left(\sum_k w^{(2)}_{kj} \left(x_{i'}\, h'(a_{j'}) \sum_{k'} M_{kk'}\, w^{(2)}_{k'j'} \right) \right)$$

$$= x_i\, x_{i'} I_{jj'} h''(a_j) \left(\sum_k w^{(2)}_{kj}\, \delta_k \right) + x_i\, x_{i'}\, h'(a_j)\, h'(a_{j'}) \sum_k \sum_{k'} w^{(2)}_{kj}\, w^{(2)}_{k'j'} M_{kk'}$$

3. One weight in each layer:

Using 5.50 and 5.56:

$$\frac{\partial E_n}{\partial w^{(1)}_{ji}} = x_i\, h'(a_j) \sum_{k'} w^{(2)}_{k'j} \delta_{k'}$$

$$\implies \frac{\partial^2 E_n}{\partial w^{(2)}_{kj'}\, \partial w^{(1)}_{ji}} = \frac{\partial}{\partial w^{(2)}_{kj'}} \left(x_i\, h'(a_j) \sum_{k'} w^{(2)}_{k'j} \delta_{k'} \right)$$

$$= \left(x_i\, h'(a_j) \sum_{k'} \left(\frac{\partial w^{(2)}_{k'j} \delta_{k'}}{\partial w^{(2)}_{kj'}} \right) \right)$$

$$= \left(x_i\, h'(a_j) \sum_{k'} \left(\delta_{k'} \frac{\partial w^{(2)}_{k'j}}{\partial w^{(2)}_{kj'}} + w^{(2)}_{k'j} \frac{\partial \delta_{k'}}{\partial w^{(2)}_{kj'}} \right) \right)$$

The derivative in the first term inside the summation can be simplified as:

$$\frac{\partial w^{(2)}_{k'j}}{\partial w^{(2)}_{kj'}} = I_{jj'}\; I_{kk'}$$

The derivative in the second term inside the summation can be simplified as:

$$\frac{\partial \delta_{k'}}{\partial w^{(2)}_{kj'}} = \frac{\partial^2 E_n}{\partial w^{(2)}_{kj'}\; \partial a_{k'}}$$

$$= \frac{\partial^2 E_n}{\partial a_k\; \partial a_{k'}}\; \frac{\partial a_k}{\partial w^{(2)}_{kj'}}$$

$$= M_{kk'}\, z_{j'}$$

Substituting the two results back, we get:

$$\Longrightarrow \frac{\partial^2 E_n}{\partial w^{(2)}_{kj'}\; \partial w^{(1)}_{ji}} = \left(x_i\, h'(a_j) \sum_{k'} \left(\delta_{k'} I_{jj'} I_{kk'} + w^{(2)}_{k'j} M_{kk'}\, z_{j'} \right) \right)$$

$$= x_i\, h'(a_j) \left(\sum_{k'} \delta_{k'} I_{jj'} I_{kk'} + \sum_{k'} w^{(2)}_{k'j} M_{kk'}\, z_{j'} \right)$$

$$= x_i\, h'(a_j) \left(I_{jj'} \sum_{k'} \delta_{k'} I_{kk'} + z_{j'} \sum_{k'} w^{(2)}_{k'j} M_{kk'} \right)$$

$$= x_i\, h'(a_j) \left(I_{jj'} \delta_k + z_{j'} \sum_{k'} w^{(2)}_{k'j} M_{kk'} \right)$$