

5.6 5.21 is the error function in binary classification. So there is no k here, as there is only 1 output unit.

$$\begin{aligned}
\frac{\partial E}{\partial a} &= \frac{\partial}{\partial a} (-\{t_n \ln y_n + (1 - t_n) \ln[1 - y_n]\}) \\
&= -\left\{t_n \left(\frac{\partial \ln y_n}{\partial a}\right) + (1 - t_n) \left(\frac{\partial \ln[1 - y_n]}{\partial a}\right)\right\} \\
&= -\left\{t_n \left(\frac{1}{y}\right) \left(\frac{\partial y_n}{\partial a}\right) + (1 - t_n) \left(\frac{1}{[1 - y_n]}\right) \left(\frac{\partial [1 - y_n]}{\partial a}\right)\right\} \\
&= -\left\{t_n \left(\frac{1}{y_n}\right) \left(\frac{\partial \sigma(a)}{\partial a}\right) + (1 - t_n) \left(\frac{1}{[1 - y_n]}\right) \left(\frac{\partial [1 - \sigma(a)]}{\partial a}\right)\right\} \\
&= -\left\{t_n \left(\frac{1}{y_n}\right) (\sigma(a)(1 - \sigma(a))) + (1 - t_n) \left(\frac{1}{[1 - y_n]}\right) (0 - \sigma(a)(1 - \sigma(a)))\right\} \\
&= -\left\{t_n \left(\frac{1}{y_n}\right) (y_n(1 - y_n)) - (1 - t_n) \left(\frac{1}{[1 - y_n]}\right) (y_n(1 - y_n))\right\} \\
&= -\{t_n(1 - y_n) - (1 - t_n)y_n\} \\
&= -\{t_n - t_n y_n - y_n + y_n t_n\} \\
&= -\{t_n - y_n\} \\
&= y_n - t_n
\end{aligned}$$

which satisfies 5.18.