5.5 The likelihood function for a multi-class neural network model is given by:

$$p(\mathbf{T}|\mathbf{X}, \mathbf{w}) = \prod_{n=1}^{N} p(\mathbf{t}_n|\mathbf{x}_n, \mathbf{W})$$

Using 5.22, this becomes:

$$= \prod_{n=1}^{N} \prod_{k=1}^{K} y_k(\mathbf{x}_n, \mathbf{w})^{t_k} [1 - y_k(\mathbf{x}_n, \mathbf{w})]^{1-t_k}$$

Minimizing the negative-log of this function gives the same result as maximizing it, givins us:

$$-\ln p(\mathbf{T}|\mathbf{X}, \mathbf{w}) = -\sum_{n=1}^{N} \sum_{k=1}^{K} \{t_k \ln y_k(\mathbf{x}_n, \mathbf{w}) + (1 - t_k) \ln[1 - y_k(\mathbf{x}_n, \mathbf{w})]\}$$

1