5.2 For a K-output neural network where the conditional distribution of the target values is given by 5.16:

$$p(\mathbf{t}|\mathbf{x}, \mathbf{w}) = \mathcal{N}\left(\mathbf{t}|\mathbf{y}(\mathbf{x}, \mathbf{W}), \beta^{-1}\mathbf{I}\right)$$

The likelihood function is given by:

$$p(\mathbf{T}|\mathbf{X}, \mathbf{W}) = \prod_{i=1}^{N} \mathcal{N}\left(\mathbf{t}_i|\mathbf{y}(\mathbf{x}_i, \mathbf{W}), \beta^{-1}\mathbf{I}\right)$$

The log-likelihood is given by:

$$\ln p(\mathbf{T}|\mathbf{X}, \mathbf{W}) = \sum_{i=1}^{N} \ln \mathcal{N}\left(\mathbf{t}_i|\mathbf{y}(\mathbf{x}_i, \mathbf{W}), \beta^{-1}\mathbf{I}\right)$$

$$= \sum_{i=1}^{N} \ln \left(\frac{1}{(2\pi)^{K/2}|\beta^{-1}\mathbf{I}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{t}_i - \mathbf{y}(\mathbf{x}_i, \mathbf{W}))^T \beta \mathbf{I}(\mathbf{t}_i - \mathbf{y}(\mathbf{x}_i, \mathbf{W}))\right\}\right)$$

$$= \sum_{i=1}^{N} \ln \left(\frac{1}{(2\pi)^{K/2}\beta^{-K/2}} \exp\left\{-\frac{\beta}{2}(\mathbf{t}_i - \mathbf{y}(\mathbf{x}_i, \mathbf{W}))^T(\mathbf{t}_i - \mathbf{y}(\mathbf{x}_i, \mathbf{W}))\right\}\right)$$

$$= \sum_{i=1}^{N} \left(\frac{K}{2}\ln\left(\frac{\beta}{2\pi}\right) - \frac{\beta}{2}(\mathbf{t}_i - \mathbf{y}(\mathbf{x}_i, \mathbf{W}))^T(\mathbf{t}_i - \mathbf{y}(\mathbf{x}_i, \mathbf{W}))\right)$$

$$= \frac{NK}{2}\ln\left(\frac{\beta}{2\pi}\right) - \frac{\beta}{2}\sum_{i=1}^{N}(\mathbf{t}_i - \mathbf{y}(\mathbf{x}_i, \mathbf{W}))^T(\mathbf{t}_i - \mathbf{y}(\mathbf{x}_i, \mathbf{W}))$$

$$= \frac{NK}{2}\ln\left(\frac{\beta}{2\pi}\right) - \frac{\beta}{2}\sum_{i=1}^{N}||\mathbf{t}_i - \mathbf{y}(\mathbf{x}_i, \mathbf{W})||^2$$

Maximizing this expression w.r.t $\mathbf{W}$ gives us:

$$\mathbf{W}_{ML} = \arg\max_{\mathbf{W}} \left(-\frac{\beta}{2}\sum_{i=1}^{N}||\mathbf{t}_i - \mathbf{y}(\mathbf{x}_i, \mathbf{W})||^2\right)$$

$$= \arg\min_{\mathbf{W}} \left(\sum_{i=1}^{N}||\mathbf{t}_i - \mathbf{y}(\mathbf{x}_i, \mathbf{W})||^2\right)$$

which is equivalent to minimizing the sum-of-squares error function (5.11).