3.6 The expression for likelihood function (similar to 3.10) is given by:

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

where

$$\mathbf{T} = \begin{bmatrix} \dots & \mathbf{t}_1 & \dots \\ \dots & \mathbf{t}_2 & \dots \\ \vdots & \vdots & \vdots \\ \dots & \mathbf{t}_N & \dots \end{bmatrix}$$

Taking log-likelihood, we get:

$$\ln p(\mathbf{T}|\mathbf{X},\mathbf{W},\mathbf{\Sigma}) = \sum_{n=1}^{N} \ln \mathcal{N}(\mathbf{t}_n|\mathbf{W}^T\phi(\mathbf{x}_n),\mathbf{\Sigma})$$

$$= \sum_{n=1}^{N} \ln\left(\frac{1}{(2\pi)^{D/2}|\mathbf{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}\left(\mathbf{t}_n - \mathbf{W}^T\phi(\mathbf{x}_n)\right)^T \mathbf{\Sigma}^{-1}\left(\mathbf{t}_n - \mathbf{W}^T\phi(\mathbf{x}_n)\right)\right)\right)$$

$$= \sum_{n=1}^{N} \left(-\frac{D}{2}\ln 2\pi - \frac{1}{2}\ln|\mathbf{\Sigma}| - \frac{1}{2}\left(\mathbf{t}_n - \mathbf{W}^T\phi(\mathbf{x}_n)\right)^T \mathbf{\Sigma}^{-1}\left(\mathbf{t}_n - \mathbf{W}^T\phi(\mathbf{x}_n)\right)\right)$$

Taking derivative w.r.t $\mathbf{W}^T$, the first 2 terms in the summation disappear. Applying result 88 from The Matrix Cookbook (since $\mathbf{\Sigma}^{-1}$ is symmetric), we get:

$$\frac{\partial \ln p(\mathbf{T}|\mathbf{X},\mathbf{W},\mathbf{\Sigma})}{\partial \mathbf{W}^T} = \sum_{n=1}^{N} -\frac{1}{2}\left(-2\mathbf{\Sigma}^{-1}\left(\mathbf{t}_n - \mathbf{W}^T\phi(\mathbf{x}_n)\right)\phi(\mathbf{x}_n)^T\right)$$

$$= \sum_{n=1}^{N} \mathbf{\Sigma}^{-1}\left(\mathbf{t}_n - \mathbf{W}^T\phi(\mathbf{x}_n)\right)\phi(\mathbf{x}_n)^T$$

$$\implies \left(\frac{\partial \ln p(\mathbf{T}|\mathbf{X},\mathbf{W},\mathbf{\Sigma})}{\partial \mathbf{W}}\right)^T = \sum_{n=1}^{N} \mathbf{\Sigma}^{-1}\left(\mathbf{t}_n - \mathbf{W}^T\phi(\mathbf{x}_n)\right)\phi(\mathbf{x}_n)^T$$

Setting this derivative to $\mathbf{0}$ vector, we get:

$$\sum_{n=1}^{N} \mathbf{\Sigma}^{-1}\mathbf{t}_n\phi(\mathbf{x}_n)^T = \sum_{n=1}^{N} \mathbf{\Sigma}^{-1}\mathbf{W}_{ML}^T\phi(\mathbf{x}_n)\phi(\mathbf{x}_n)^T$$

$$\implies \sum_{n=1}^{N} \mathbf{t}_n\phi(\mathbf{x}_n)^T = \sum_{n=1}^{N} \mathbf{W}_{ML}^T\phi(\mathbf{x}_n)\phi(\mathbf{x}_n)^T$$

$$\implies \sum_{n=1}^{N} \mathbf{t}_n\phi(\mathbf{x}_n)^T = \mathbf{W}_{ML}^T\left(\sum_{n=1}^{N}\phi(\mathbf{x}_n)\phi(\mathbf{x}_n)^T\right)$$

$\mathbf{\Phi}$ being defined by 3.16, we get:

$$\implies \mathbf{T}^T\mathbf{\Phi} = \mathbf{W}_{ML}^T\mathbf{\Phi}^T\mathbf{\Phi}$$

$$\implies \mathbf{\Phi}^T\mathbf{T} = \mathbf{\Phi}^T\mathbf{\Phi}\mathbf{W}_{ML}$$

$$\implies \mathbf{W}_{ML} = \left(\mathbf{\Phi}^T\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}^T\mathbf{T}$$

Here, it's easy to see that the j^{th} column of $\mathbf{W}_{ML}$ is given by $\left(\mathbf{\Phi}^T\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}^T\mathbf{t}_j$ where $\mathbf{t}_j$ is the j^{th} column of $\mathbf{T}$.

Now to find the maximum likelihood solution for $\mathbf{\Sigma}$ (same result as exercise 2.34). Taking derivative w.r.t $\mathbf{\Sigma}$, the first term in the summation disappears. Applying results 57 and 61 from The Matrix Cookbook (since $\mathbf{\Sigma}^{-1}$ is symmetric), we get:

$$\frac{\partial \ln p(\mathbf{T}|\mathbf{X}, \mathbf{W}, \mathbf{\Sigma})}{\partial \mathbf{\Sigma}} = \sum_{n=1}^{N}\left(-\frac{1}{2}\mathbf{\Sigma}^{-T} - \frac{1}{2}\left(-\mathbf{\Sigma}^{-T}\left(\mathbf{t}_n - \mathbf{W}^T\phi(\mathbf{x}_n)\right)\left(\mathbf{t}_n - \mathbf{W}^T\phi(\mathbf{x}_n)\right)^T\mathbf{\Sigma}^{-T}\right)\right)$$

$$= \sum_{n=1}^{N}\left(-\frac{1}{2}\mathbf{\Sigma}^{-1} + \frac{1}{2}\mathbf{\Sigma}^{-1}\left(\mathbf{t}_n - \mathbf{W}^T\phi(\mathbf{x}_n)\right)\left(\mathbf{t}_n - \mathbf{W}^T\phi(\mathbf{x}_n)\right)^T\mathbf{\Sigma}^{-1}\right)$$

$$= -\frac{N}{2}\mathbf{\Sigma}^{-1} + \sum_{n=1}^{N}\frac{1}{2}\mathbf{\Sigma}^{-1}\left(\mathbf{t}_n - \mathbf{W}^T\phi(\mathbf{x}_n)\right)\left(\mathbf{t}_n - \mathbf{W}^T\phi(\mathbf{x}_n)\right)^T\mathbf{\Sigma}^{-1}$$

Setting this derivative to $\mathbf{0}$, we get:

$$
\begin{aligned}
\frac{N}{2}\boldsymbol{\Sigma}^{-1} &= \sum_{n=1}^{N} \frac{1}{2}\boldsymbol{\Sigma}^{-1}\left(\mathbf{t}_n - \mathbf{W}^T\phi(\mathbf{x}_n)\right)\left(\mathbf{t}_n - \mathbf{W}^T\phi(\mathbf{x}_n)\right)^T\boldsymbol{\Sigma}^{-1} \\
\implies \frac{N}{2} &= \sum_{n=1}^{N} \frac{1}{2}\boldsymbol{\Sigma}^{-1}\left(\mathbf{t}_n - \mathbf{W}^T\phi(\mathbf{x}_n)\right)\left(\mathbf{t}_n - \mathbf{W}^T\phi(\mathbf{x}_n)\right)^T \\
\implies N\boldsymbol{\Sigma} &= \sum_{n=1}^{N} \left(\mathbf{t}_n - \mathbf{W}^T\phi(\mathbf{x}_n)\right)\left(\mathbf{t}_n - \mathbf{W}^T\phi(\mathbf{x}_n)\right)^T \\
\implies \boldsymbol{\Sigma} &= \frac{1}{N}\sum_{n=1}^{N} \left(\mathbf{t}_n - \mathbf{W}^T\phi(\mathbf{x}_n)\right)\left(\mathbf{t}_n - \mathbf{W}^T\phi(\mathbf{x}_n)\right)^T
\end{aligned}
$$

We can substitute $\mathbf{W}_{ML}$ for $\mathbf{W}$ here to get the maximum likelihood solution:

$$
\implies \boldsymbol{\Sigma}_{ML} = \frac{1}{N}\sum_{n=1}^{N} \left(\mathbf{t}_n - \mathbf{W}_{ML}^T\phi(\mathbf{x}_n)\right)\left(\mathbf{t}_n - \mathbf{W}_{ML}^T\phi(\mathbf{x}_n)\right)^T.
$$