3.7 The posterior distribution is given by:

$$p(\mathbf{w}|\mathbf{t}) \propto p(\mathbf{t}|\mathbf{X}, \mathbf{w}, \beta) * p(\mathbf{w})$$

From 3.10, and 3.48, this becomes:

$$= \left(\prod_{n=1}^{N} \mathcal{N}\left(\mathbf{t}_n | \mathbf{w}^T \phi(\mathbf{x}_n), \beta^{-1}\right) \right) \mathcal{N}(\mathbf{w}|\mathbf{m}_0, \mathbf{S}_0)$$

We can consider one n at a time, and apply 2.116 to obtain the posterior, such that:

$$p(\mathbf{w}|t_1) = \mathcal{N}\left(t_1 | \mathbf{w}^T \phi(\mathbf{x}_1), \beta^{-1}\right) \mathcal{N}(\mathbf{w}|\mathbf{m}_0, \mathbf{S}_0)$$

Here, comparing the R.H.S TO 2.113 and 2.114, we see that:

$$\begin{aligned} \mathbf{x} &= \mathbf{w} \\ \boldsymbol{\mu} &= \mathbf{m}_0 \\ \boldsymbol{\Lambda} &= \mathbf{S}_0^{-1} \\ \\ \mathbf{y} &= t_1 \\ \mathbf{A} &= \phi(\mathbf{x}_1)^T \\ b &= 0 \\ \mathbf{L} &= \beta \end{aligned}$$

Applying the result from 2.116, we get:

$$= \mathcal{N}\left(\mathbf{w} | \mathbf{S}_1 \{\phi(\mathbf{x}_1)\beta(t_1) + \mathbf{S}_0^{-1}\mathbf{m}_0\}, \mathbf{S}_1\right)$$

where $\mathbf{S}_1 = (\mathbf{S}_0^{-1} + \phi(\mathbf{x}_1)\beta\phi(\mathbf{x}_1)^T)^{-1}$

Similarly, we consider the next data point and calculate the posterior:

$$p(\mathbf{w}|t_1, t_2) = \mathcal{N}\left(t_2 | \mathbf{w}^T \phi(\mathbf{x}_2), \beta^{-1}\right) \mathcal{N}\left(\mathbf{w} | \mathbf{S}_1 \{\phi(\mathbf{x}_1)\beta t_1 + \mathbf{S}_0^{-1}\mathbf{m}_0\}, \mathbf{S}_1\right)$$

Here, comparing the R.H.S TO 2.113 and 2.114, we see that:

$$\mathbf{x} = \mathbf{w}$$
$$\boldsymbol{\mu} = \mathbf{S}_1\{\phi(\mathbf{x}_1)\beta t_1 + \mathbf{S}_0^{-1}\mathbf{m}_0\}$$
$$\boldsymbol{\Lambda} = \mathbf{S}_1^{-1}$$

$$\mathbf{y} = t_2$$
$$\mathbf{A} = \phi(\mathbf{x}_2)^T$$
$$b = 0$$
$$\mathbf{L} = \beta$$

Applying the result from 2.116, we get:

$$= \mathcal{N}\left(\mathbf{w}|\mathbf{S}_2\{\phi(\mathbf{x}_2)\beta t_2 + \mathbf{S}_1^{-1}\mathbf{S}_1\{\phi(\mathbf{x}_1)\beta t_1 + \mathbf{S}_0^{-1}\mathbf{m}_0\}\}, \mathbf{S}_2\right)$$

$$= \mathcal{N}\left(\mathbf{w}|\mathbf{S}_2\{\phi(\mathbf{x}_2)\beta t_2 + \{\phi(\mathbf{x}_1)\beta t_1 + \mathbf{S}_0^{-1}\mathbf{m}_0\}\}, \mathbf{S}_2\right)$$

$$= \mathcal{N}\left(\mathbf{w}|\mathbf{S}_2\{\beta(\phi(\mathbf{x}_2)t_2 + \phi(\mathbf{x}_1)t_1) + \mathbf{S}_0^{-1}\mathbf{m}_0\}, \mathbf{S}_2\right)$$

$$= \mathcal{N}\left(\mathbf{w}|\mathbf{S}_2\{\beta(\phi(\mathbf{x}_2)t_2 + \phi(\mathbf{x}_1)t_1) + \mathbf{S}_0^{-1}\mathbf{m}_0\}, \mathbf{S}_2\right)$$

where $\mathbf{S}_2 = \left(\mathbf{S}_1^{-1} + \phi(\mathbf{x}_2)\beta\phi(\mathbf{x}_2)^T\right)^{-1} = \left(\mathbf{S}_0^{-1} + \phi(\mathbf{x}_1)\beta\phi(\mathbf{x}_1)^T + \phi(\mathbf{x}_2)\beta\phi(\mathbf{x}_2)^T\right)^{-1}$

As we can see, the term $\beta\phi(\mathbf{x}_n)\beta\phi(\mathbf{x}_n)^T$ gets added to the expression for covariance with each successive data point. Therefore:

$$\mathbf{S}_N = \left(\mathbf{S}_0^{-1} + \phi(\mathbf{x}_1)\beta\phi(\mathbf{x}_1)^T + \phi(\mathbf{x}_2)\beta\phi(\mathbf{x}_2)^T + \ldots + \phi(\mathbf{x}_N)\beta\phi(\mathbf{x}_N)^T\right)^{-1}$$

$$= \left(\mathbf{S}_0^{-1} + \beta\sum_{n=1}^{N}\phi(\mathbf{x}_n)\phi(\mathbf{x}_n)^T\right)^{-1}$$
$$\implies \mathbf{S}_N = \left(\mathbf{S}_0^{-1} + \beta\boldsymbol{\Phi}^T\boldsymbol{\Phi}\right)^{-1}$$

which is the same as 3.51.

Similarly, in the expression of the mean, $\beta\phi(\mathbf{x}_n)t_n$ is getting added inside the curly braces with each successive data point. Therefore:

$$\begin{aligned}\mathbf{m}_N &= \mathbf{S}_N \left\{ \beta \left(\sum_{n=1}^{N} \phi(\mathbf{x}_n)t_n \right) + \mathbf{S}_0^{-1}\mathbf{m}_0 \right\} \\ &= \mathbf{S}_N \left\{ \beta\boldsymbol{\Phi}^T\mathbf{t} + \mathbf{S}_0^{-1}\mathbf{m}_0 \right\}\end{aligned}$$

which is the same as 3.50.