

**3.9** 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:

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

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

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

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

$$\mathbf{x} = \mathbf{w}$$

$$\boldsymbol{\mu} = \mathbf{m}_0$$

$$\boldsymbol{\Lambda} = \mathbf{S}_0^{-1}$$

$$\mathbf{y} = t_1$$

$$\mathbf{A} = \phi(\mathbf{x}_1)^T$$

$$b = 0$$

$$\mathbf{L} = \beta$$

Applying the result from 2.116, we get:

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

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:

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

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

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

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

Applying the result from 2.116, we get:

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

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

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:

$$\begin{aligned}\mathbf{S}_N &= (\mathbf{S}_0^{-1} + \phi(\mathbf{x}_1)\beta\phi(\mathbf{x}_1)^T + \phi(\mathbf{x}_2)\beta\phi(\mathbf{x}_2)^T + \dots + \phi(\mathbf{x}_N)\beta\phi(\mathbf{x}_N)^T)^{-1} \\ &= \left(\mathbf{S}_0^{-1} + \beta \sum_{n=1}^N \phi(\mathbf{x}_n)\phi(\mathbf{x}_n)^T\right)^{-1} \\ &\implies \mathbf{S}_N = (\mathbf{S}_0^{-1} + \beta\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\end{aligned}$$

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:

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

which is the same as 3.50.

Now we get the result in exercise 3.8 by using 2.116:

$$p(\mathbf{w}|t_{N+1}, \mathbf{t}_N) \propto p(t_{N+1}|\mathbf{w}^T\phi(\mathbf{x}_{N+1}), \beta^{-1}) * p(\mathbf{w}|\mathbf{m}_N, \mathbf{S}_N)$$

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

$$\begin{aligned}\mathbf{x} &= \mathbf{w} \\ \boldsymbol{\mu} &= \mathbf{m}_N \\ \boldsymbol{\Lambda} &= \mathbf{S}_N^{-1} \\ \mathbf{y} &= t_{N+1} \\ \mathbf{A} &= \phi(\mathbf{x}_{N+1})^T \\ b &= 0 \\ \mathbf{L} &= \beta\end{aligned}$$

Applying the result from 2.116, we get:

$$\begin{aligned}p(\mathbf{w}|t_{N+1}, \mathbf{t}_N) &= \mathcal{N}(\mathbf{w}|\mathbf{S}_{N+1}\{\phi(\mathbf{x}_{N+1})\beta t_{N+1} + \mathbf{S}_N^{-1}\mathbf{m}_N\}, \mathbf{S}_{N+1}) \\ &= \mathcal{N}\left(\mathbf{w}|\mathbf{S}_{N+1} \left\{ \phi(\mathbf{x}_{N+1})\beta t_{N+1} + \beta \left( \sum_{n=1}^N \phi(\mathbf{x}_n)t_n \right) + \mathbf{S}_0^{-1}\mathbf{m}_0 \right\}, \mathbf{S}_{N+1}\right)\end{aligned}$$

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