

**3.12** The posterior distribution is given by:

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

where the R.H.S terms are given by 3.10 and 3.112:

$$p(\mathbf{t} | \mathbf{X}, \mathbf{w}, \beta) = \prod_{n=1}^N p(t_n | \mathbf{x}_n, \mathbf{w}, \beta) = \prod_{n=1}^N \mathcal{N}(t_n | \mathbf{w}^T \phi(\mathbf{x}_n), \beta^{-1})$$

$$p(\mathbf{w}, \beta) = \mathcal{N}(\mathbf{w} | \mathbf{m}_0, \beta^{-1} \mathbf{S}_0) \text{Gam}(\beta | a_0, b_0)$$

Given the above, R.H.S is given by:

$$\begin{aligned} & \left( \prod_{n=1}^N \mathcal{N}(t_n | \mathbf{w}^T \phi(\mathbf{x}_n), \beta^{-1}) \right) \mathcal{N}(\mathbf{w} | \mathbf{m}_0, \beta^{-1} \mathbf{S}_0) \text{Gam}(\beta | a_0, b_0) \\ &= \left( \prod_{n=1}^N \frac{1}{\sqrt{2\pi\beta^{-1}}} \exp \left\{ -\frac{\beta(t_n - \mathbf{w}^T \phi(\mathbf{x}_n))^2}{2} \right\} \right) \\ & \frac{1}{(2\pi)^{D/2} |\beta^{-1} \mathbf{S}_0|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{w} - \mathbf{m}_0)^T \beta \mathbf{S}_0^{-1} (\mathbf{w} - \mathbf{m}_0) \right\} \\ & \frac{1}{\Gamma(a_0)} b_0^{a_0} \beta^{a_0-1} \exp(-b_0 \beta) \\ &= \frac{1}{(2\pi\beta^{-1})^{N/2}} \exp \left\{ \sum_{n=1}^N -\frac{\beta(t_n - \mathbf{w}^T \phi(\mathbf{x}_n))^2}{2} \right\} \\ & \frac{1}{(2\pi)^{D/2} |\beta^{-1} \mathbf{S}_0|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{w} - \mathbf{m}_0)^T \beta \mathbf{S}_0^{-1} (\mathbf{w} - \mathbf{m}_0) \right\} \\ & \frac{1}{\Gamma(a_0)} b_0^{a_0} \beta^{a_0-1} \exp(-b_0 \beta) \\ &= \frac{\beta^{(N+1)/2}}{(2\pi)^{(N+D)/2} |\mathbf{S}_0|^{1/2}} \exp \left\{ \sum_{n=1}^N -\frac{\beta(t_n - \mathbf{w}^T \phi(\mathbf{x}_n))^2}{2} \right\} \\ & \exp \left\{ -\frac{1}{2} (\mathbf{w} - \mathbf{m}_0)^T \beta \mathbf{S}_0^{-1} (\mathbf{w} - \mathbf{m}_0) \right\} \end{aligned}$$

$$\frac{1}{\Gamma(a_0)} b_0^{a_0} \beta^{a_0-1} \exp(-b_0 \beta)$$

First we simplify the exponent of the first term. From the solution to exercise 3.7, we know that it becomes:

$$\begin{aligned} &= \frac{\beta^{(N+1)/2}}{(2\pi)^{(N+D)/2} |\mathbf{S}_0|^{1/2}} \exp \left\{ -\frac{\beta}{2} \left( \mathbf{t}^T \mathbf{t} - 2\mathbf{w}^T \boldsymbol{\Phi}^T \mathbf{t} + \mathbf{w}^T \boldsymbol{\Phi}^T \boldsymbol{\Phi} \mathbf{w} \right) \right\} \\ &\quad \exp \left\{ -\frac{\beta}{2} (\mathbf{w} - \mathbf{m}_0)^T \mathbf{S}_0^{-1} (\mathbf{w} - \mathbf{m}_0) \right\} \\ &\quad \frac{1}{\Gamma(a_0)} b_0^{a_0} \beta^{a_0-1} \exp(-b_0 \beta) \end{aligned}$$

Expanding the exponent of the second term and merging with the first, we get:

$$\begin{aligned} &-\frac{\beta}{2} \left( \mathbf{t}^T \mathbf{t} - 2\mathbf{w}^T \boldsymbol{\Phi}^T \mathbf{t} + \mathbf{w}^T \boldsymbol{\Phi}^T \boldsymbol{\Phi} \mathbf{w} + \mathbf{w}^T \mathbf{S}_0^{-1} \mathbf{w} - 2\mathbf{w}^T \mathbf{S}_0^{-1} \mathbf{m}_0 + \mathbf{m}_0^T \mathbf{S}_0^{-1} \mathbf{m}_0 \right) \\ &= -\frac{\beta}{2} \left( \mathbf{w}^T (\boldsymbol{\Phi}^T \boldsymbol{\Phi} + \mathbf{S}_0^{-1}) \mathbf{w} - 2\mathbf{w}^T (\boldsymbol{\Phi}^T \mathbf{t} + \mathbf{S}_0^{-1} \mathbf{m}_0) + \mathbf{m}_0^T \mathbf{S}_0^{-1} \mathbf{m}_0 + \mathbf{t}^T \mathbf{t} \right) \\ &= -\frac{\beta}{2} \left( \mathbf{w}^T \mathbf{S}_N^{-1} \mathbf{w} - 2\mathbf{w}^T \mathbf{S}_N^{-1} \mathbf{m}_N + \mathbf{m}_0^T \mathbf{S}_0^{-1} \mathbf{m}_0 + \mathbf{t}^T \mathbf{t} \right) \end{aligned}$$

Adding and subtracting  $\mathbf{m}_N^T \mathbf{S}_N^{-1} \mathbf{m}_N$ , we get:

$$\begin{aligned} &= -\frac{\beta}{2} \left( \mathbf{w}^T \mathbf{S}_N^{-1} \mathbf{w} - 2\mathbf{w}^T \mathbf{S}_N^{-1} \mathbf{m}_N + \mathbf{m}_N^T \mathbf{S}_N^{-1} \mathbf{m}_N - \mathbf{m}_N^T \mathbf{S}_N^{-1} \mathbf{m}_N + \mathbf{m}_0^T \mathbf{S}_0^{-1} \mathbf{m}_0 + \mathbf{t}^T \mathbf{t} \right) \\ &= -\frac{\beta}{2} \left( (\mathbf{w} - \mathbf{m}_N)^T \mathbf{S}_N^{-1} (\mathbf{w} - \mathbf{m}_N) - \mathbf{m}_N^T \mathbf{S}_N^{-1} \mathbf{m}_N + \mathbf{m}_0^T \mathbf{S}_0^{-1} \mathbf{m}_0 + \mathbf{t}^T \mathbf{t} \right) \end{aligned}$$

Substituting this result back, we get:

$$= \frac{\beta^{(N+1)/2}}{(2\pi)^{(N+D)/2} |\mathbf{S}_0|^{1/2}}$$

$$\begin{aligned}
& \exp \left\{ -\frac{\beta}{2} ((\mathbf{w} - \mathbf{m}_N)^T \mathbf{S}_N^{-1} (\mathbf{w} - \mathbf{m}_N) - \mathbf{m}_N^T \mathbf{S}_N^{-1} \mathbf{m}_N + \mathbf{m}_0^T \mathbf{S}_0^{-1} \mathbf{m}_0 + \mathbf{t}^T \mathbf{t}) \right\} \\
& \frac{1}{\Gamma(a_0)} b_0^{a_0} \beta^{a_0-1} \exp(-b_0 \beta) \\
& = \frac{\beta^{(N+1)/2}}{(2\pi)^{(N+D)/2} |\mathbf{S}_0|^{1/2}} \\
& \exp \left\{ -\frac{\beta}{2} ((\mathbf{w} - \mathbf{m}_N)^T \mathbf{S}_N^{-1} (\mathbf{w} - \mathbf{m}_N)) \right\} \\
& \exp \left\{ -\frac{\beta}{2} (-\mathbf{m}_N^T \mathbf{S}_N^{-1} \mathbf{m}_N + \mathbf{m}_0^T \mathbf{S}_0^{-1} \mathbf{m}_0 + \mathbf{t}^T \mathbf{t}) \right\} \\
& \frac{1}{\Gamma(a_0)} b_0^{a_0} \beta^{a_0-1} \exp(-b_0 \beta) \\
& = \left( \frac{1}{(2\pi)^{D/2} |\beta^{-1} \mathbf{S}_N|^{1/2}} \right) \left( \frac{\beta^{N/2} |\mathbf{S}_N|^{1/2}}{(2\pi)^{N/2} |\mathbf{S}_0|^{1/2}} \right) \\
& \exp \left\{ -\frac{\beta}{2} ((\mathbf{w} - \mathbf{m}_N)^T \mathbf{S}_N^{-1} (\mathbf{w} - \mathbf{m}_N)) \right\} \\
& \exp \left\{ -\frac{\beta}{2} (-\mathbf{m}_N^T \mathbf{S}_N^{-1} \mathbf{m}_N + \mathbf{m}_0^T \mathbf{S}_0^{-1} \mathbf{m}_0 + \mathbf{t}^T \mathbf{t}) \right\} \\
& \frac{1}{\Gamma(a_0)} b_0^{a_0} \beta^{a_0-1} \exp(-b_0 \beta) \\
& = \left( \frac{1}{(2\pi)^{D/2} |\beta^{-1} \mathbf{S}_N|^{1/2}} \right) \exp \left\{ -\frac{\beta}{2} ((\mathbf{w} - \mathbf{m}_N)^T (\beta^{-1} \mathbf{S}_N)^{-1} (\mathbf{w} - \mathbf{m}_N)) \right\} \\
& \left( \frac{\beta^{N/2} |\mathbf{S}_N|^{1/2}}{(2\pi)^{N/2} |\mathbf{S}_0|^{1/2}} \right) \exp \left\{ -\frac{\beta}{2} (-\mathbf{m}_N^T \mathbf{S}_N^{-1} \mathbf{m}_N + \mathbf{m}_0^T \mathbf{S}_0^{-1} \mathbf{m}_0 + \mathbf{t}^T \mathbf{t}) \right\} \\
& \frac{1}{\Gamma(a_0)} b_0^{a_0} \beta^{a_0-1} \exp(-b_0 \beta) \\
& = \mathcal{N}(\mathbf{w} | \mathbf{m}_N, \beta^{-1} \mathbf{S}_N) \\
& \left( \frac{\beta^{N/2} |\mathbf{S}_N|^{1/2}}{(2\pi)^{N/2} |\mathbf{S}_0|^{1/2}} \right) \exp \left\{ -\frac{\beta}{2} (-\mathbf{m}_N^T \mathbf{S}_N^{-1} \mathbf{m}_N + \mathbf{m}_0^T \mathbf{S}_0^{-1} \mathbf{m}_0 + \mathbf{t}^T \mathbf{t}) \right\}
\end{aligned}$$

$$\begin{aligned}
& \frac{1}{\Gamma(a_0)} b_0^{a_0} \beta^{a_0-1} \exp(-b_0 \beta) \\
&= \mathcal{N}(\mathbf{w}|\mathbf{m}_N, \beta^{-1} \mathbf{S}_N) \\
& \left( \frac{|\mathbf{S}_N|^{1/2}}{(2\pi)^{N/2} |\mathbf{S}_0|^{1/2}} \right) \frac{1}{\Gamma(a_0)} b_0^{a_0} \beta^{(a_0-1+N/2)} \\
& \exp \left\{ -\beta \left( \frac{1}{2} (\mathbf{m}_N^T \mathbf{S}_N^{-1} \mathbf{m}_N + \mathbf{m}_0^T \mathbf{S}_0^{-1} \mathbf{m}_0 + \mathbf{t}^T \mathbf{t}) + b_0 \right) \right\}
\end{aligned}$$

Here, let  $a_N = a_0 + N/2$ , and  $b_N = b_0 + \frac{1}{2} (\mathbf{m}_N^T \mathbf{S}_N^{-1} \mathbf{m}_N + \mathbf{m}_0^T \mathbf{S}_0^{-1} \mathbf{m}_0 + \mathbf{t}^T \mathbf{t})$ .

The expression becomes:

$$\begin{aligned}
&= \mathcal{N}(\mathbf{w}|\mathbf{m}_N, \beta^{-1} \mathbf{S}_N) \\
& \left( \frac{|\mathbf{S}_N|^{1/2}}{(2\pi)^{N/2} |\mathbf{S}_0|^{1/2}} \right) \frac{1}{\Gamma(a_0)} b_0^{a_0} \beta^{(a_N-1)} \exp \{-\beta b_N\} \\
&= \mathcal{N}(\mathbf{w}|\mathbf{m}_N, \beta^{-1} \mathbf{S}_N) \\
& \left( \frac{|\mathbf{S}_N|^{1/2}}{(2\pi)^{N/2} |\mathbf{S}_0|^{1/2}} \frac{\Gamma(a_N)}{\Gamma(a_0)} \frac{b_0^{a_0}}{b_N^{a_N}} \right) \\
& \frac{1}{\Gamma(a_N)} b_N^{a_N} \beta^{(a_N-1)} \exp \{-\beta b_N\}
\end{aligned}$$

PS: This type of multiplying and dividing by the term  $\frac{1}{\Gamma(a_N)} b_N^{a_N}$  can be done because the Gamma distribution is a distribution of  $\beta$ , and this term is just the normalizing constant.

This finally gives us:

$$p(\mathbf{w}, \beta | \mathbf{t}) = c \mathcal{N}(\mathbf{w}|\mathbf{m}_N, \beta^{-1} \mathbf{S}_N) \text{Gam}(\beta | a_N, b_N)$$

$$\text{where } c = \left( \frac{|\mathbf{S}_N|^{1/2}}{(2\pi)^{N/2} |\mathbf{S}_0|^{1/2}} \frac{\Gamma(a_N)}{\Gamma(a_0)} \frac{b_0^{a_0}}{b_N^{a_N}} \right)$$