EECE7397 – Homework 3

1.

$$E_D(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N} r_n \left\{t_n - \mathbf{w}^T\phi(\mathbf{x}_n)\right\}^2$$

$$\begin{aligned}
\nabla_{\mathbf{w}} E_D(\mathbf{w}) &= \frac{\partial}{\partial \mathbf{w}}\frac{1}{2}\sum_{n=1}^{N} r_n \left\{t_n - \mathbf{w}^T\phi(\mathbf{x}_n)\right\}^2 \\
&= \frac{1}{2}\sum_{n=1}^{N} r_n \cdot 2\left\{t_n - \mathbf{w}^T\phi(\mathbf{x}_n)\right\}(-\phi(\mathbf{x}_n)^T) \\
&= -\sum_{n=1}^{N} r_n \left\{\mathbf{w}^T\phi(\mathbf{x}_n) - t_n\right\}\phi(\mathbf{x}_n)^T \\
&= -\sum_{n=1}^{N} r_n \left\{\mathbf{w}^T\phi(\mathbf{x}_n)\phi(\mathbf{x}_n)^T - t_n\phi(\mathbf{x}_n)^T\right\} \\
&= \sum_{n=1}^{N} r_n t_n \phi(\mathbf{x}_n)^T - \mathbf{w}^T\sum_{n=1}^{N} r_n \phi(\mathbf{x}_n)\phi(\mathbf{x}_n)^T
\end{aligned}$$

$$\begin{aligned}
&\nabla_{\mathbf{w}} E_D(\mathbf{w}) = 0 \\
&\Rightarrow \sum_{n=1}^{N} r_n t_n \phi(\mathbf{x}_n)^T - \mathbf{w}^T\sum_{n=1}^{N} r_n \phi(\mathbf{x}_n)\phi(\mathbf{x}_n)^T = 0 \\
&\Rightarrow \mathbf{w}^T\sum_{n=1}^{N} r_n \phi(\mathbf{x}_n)\phi(\mathbf{x}_n)^T = \sum_{n=1}^{N} r_n t_n \phi(\mathbf{x}_n)^T \\
&\Rightarrow \sum_{n=1}^{N} r_n \phi(\mathbf{x}_n)\phi(\mathbf{x}_n)^T\mathbf{w} = \sum_{n=1}^{N} r_n t_n \phi(\mathbf{x}_n) \\
&\Rightarrow \Phi^T\mathbf{R}\Phi\mathbf{w} = \Phi^T\mathbf{R}\mathbf{t} \\
&\Rightarrow \mathbf{w}^* = \left(\Phi^T\mathbf{R}\Phi\right)^{-1}\Phi^T\mathbf{R}\mathbf{t}
\end{aligned}$$

For weighted sum-of-squares for replicated data points, if original data point (x_n, t_n) is replicated r_n times or the original data point is observed r_n times, then there are essentially $\sum_n r_n$ data points with the same (x_n, t_n) resulting in the same function as the original data dependent variance version.

$$\frac{1}{2}\sum_{n=1}^{N}\underbrace{\left[(t_n - \mathbf{w}^T\phi(x_n))^2 + \cdots + (t_n - \mathbf{w}^T\phi(x_n))^2\right]}_{r_n \text{ times}}$$

$$= \frac{1}{2}\sum_{n=1}^{N} r_n(t_n - \mathbf{w}^T\phi(x_n))^2$$

$$E_D(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N} r_n(t_n - \mathbf{w}^T\phi(x_n))^2.$$

Thus, minimizing the error function on the replicated data points data set (with points repeated r_n times) gives the identical solution $\mathbf{w}^*$ as the weighted sum-of-squares error function minimization on the original data.

2.

$$p(\mathbf{w} \mid \mathbf{t}) \;=\; \mathcal{N}(\mathbf{w} \mid \mathbf{m}_N, \mathbf{S}_N)$$

$$p(\mathbf{w} \mid \mathbf{t}) \;\propto\; p(\mathbf{t} \mid \mathbf{w})\, p(\mathbf{w})$$

$$p(\mathbf{t} \mid \mathbf{w}) = \prod_{n=1}^{N} \mathcal{N}\big(t_n \mid \mathbf{w}^T\phi(x_n), \beta^{-1}\big)$$

$$p(\mathbf{w}) \;=\; \mathcal{N}(\mathbf{w} \mid \mathbf{m}_0, \mathbf{S}_0)$$

$$\mathcal{N}(\mathbf{w} \mid \mathbf{m}_N, \mathbf{S}_N) \propto \prod_{n=1}^{N} \mathcal{N}\big(t_n \mid \mathbf{w}^T\phi(x_n), \beta^{-1}\big)\mathcal{N}(\mathbf{w} \mid \mathbf{m}_0, \mathbf{S}_0)$$

$$\begin{aligned}
\ln \mathcal{N}(\mathbf{w} \mid \mathbf{m}_0, \mathbf{S}_0) &= \ln\left(\frac{1}{(2\pi)^{D/2}\,|\mathbf{S}_0|^{1/2}}\exp\Big\{-\frac{1}{2}\,(\mathbf{w}-\mathbf{m}_0)^T\,\mathbf{S}_0{}^{-1}\,(\mathbf{w}-\mathbf{m}_0)\Big\}\right) \\
&= \ln\frac{1}{(2\pi)^{D/2}\,|\mathbf{S}_0|^{1/2}} + \ln\exp\Big\{-\frac{1}{2}\,(\mathbf{w}-\mathbf{m}_0)^T\,\mathbf{S}_0{}^{-1}\,(\mathbf{w}-\mathbf{m}_0)\Big\} \\
&= -\frac{D}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{S}_0| - \frac{1}{2}(\mathbf{w}-\mathbf{m}_0)^T\mathbf{S}_0^{-1}(\mathbf{w}-\mathbf{m}_0) \\
&= -\frac{1}{2}\left(\mathbf{w}^T\mathbf{S}_0^{-1}\mathbf{w} \;-\; 2\,\mathbf{w}^T\mathbf{S}_0^{-1}\mathbf{m}_0 \;+\; \mathbf{m}_0^T\mathbf{S}_0^{-1}\mathbf{m}_0\right) \\
&= -\frac{1}{2}\,\mathbf{w}^T\mathbf{S}_0^{-1}\mathbf{w} \;+\; \mathbf{w}^T\mathbf{S}_0^{-1}\mathbf{m}_0 \;-\; \frac{1}{2}\,\mathbf{m}_0^T\mathbf{S}_0^{-1}\mathbf{m}_0
\end{aligned}$$

$$
\begin{aligned}
\ln \prod_{n=1}^{N} \mathcal{N}\big(t_n \mid \mathbf{w}^T\phi(x_n),\ \beta^{-1}\big) &= \sum_{n=1}^{N} \ln \mathcal{N}\big(t_n \mid \mathbf{w}^T\phi(x_n), \beta^{-1}\big) \\
&= \sum_{n=1}^{N} \ln\left(\frac{1}{(2\pi)^{1/2}(\beta^{-1})^{1/2}} \exp\Big\{-\frac{1}{2}(t_n - \mathbf{w}^T\phi(x_n))^2\beta\Big\}\right) \\
&= \sum_{n=1}^{N} \left[\ln \frac{1}{(2\pi)^{1/2}(\beta^{-1})^{1/2}} + \ln\exp\Big\{-\frac{1}{2}(t_n - \mathbf{w}^T\phi(x_n))^2\beta\Big\}\right] \\
&= \sum_{n=1}^{N} \left[-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\beta^{-1}) - \frac{1}{2}(t_n - \mathbf{w}^T\phi(x_n))^2\beta\right] \\
&= -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\beta^{-1}) - \frac{1}{2}\beta\sum_{n=1}^{N}(t_n - \mathbf{w}^T\phi(x_n))^2 \\
&= -\frac{N}{2}\ln(2\pi) + \frac{N}{2}\ln\beta - \frac{1}{2}\beta\sum_{n=1}^{N}(t_n - \mathbf{w}^T\phi(x_n))^2 \\
&= -\frac{\beta}{2}\sum_{n=1}^{N}\big(t_n \ - \ \mathbf{w}^T\phi(x_n)\big)^2 \\
&= -\frac{\beta}{2}\sum_{n=1}^{N}\big(t_n^2 \ - \ 2\,t_n\,(\mathbf{w}^T\phi(x_n)) \ + \ (\mathbf{w}^T\phi(x_n))^2\big) \\
&= -\frac{\beta}{2}\sum_{n=1}^{N}\big(t_n^2 \ - \ 2\,t_n\,(\mathbf{w}^T\phi(x_n)) \ + \ \mathbf{w}^T\phi(x_n)\,\phi(x_n)^T\mathbf{w}\big) \\
&= -\frac{\beta}{2}\sum_{n=1}^{N} t_n^2 \ + \ \beta\sum_{n=1}^{N} t_n\,\big(\mathbf{w}^T\phi(x_n)\big) \ - \ \frac{\beta}{2}\sum_{n=1}^{N}\mathbf{w}^T\phi(x_n)\,\phi(x_n)^T\mathbf{w}
\end{aligned}
$$

$$
\begin{aligned}
\ln p(\mathbf{w} \mid \mathbf{t}) &= \ln p(\mathbf{t} \mid \mathbf{w})\, p(\mathbf{w}) \\
&= \ln\left(\prod_{n=1}^{N}\mathcal{N}\big(t_n \mid \mathbf{w}^T\phi(x_n), \beta^{-1}\big)\mathcal{N}(\mathbf{w} \mid \mathbf{m}_0, \mathbf{S}_0)\right) \\
&= \ln\prod_{n=1}^{N}\mathcal{N}\big(t_n \mid \mathbf{w}^T\phi(x_n), \beta^{-1}\big) + \ln\mathcal{N}(\mathbf{w} \mid \mathbf{m}_0, \mathbf{S}_0) \\
&= -\frac{\beta}{2}\sum_{n=1}^{N} t_n^2 \ + \ \beta\sum_{n=1}^{N} t_n\,(\mathbf{w}^T\phi(x_n)) \ - \ \frac{\beta}{2}\sum_{n=1}^{N}\mathbf{w}^T\phi(x_n)\,\phi(x_n)^T\mathbf{w} - \frac{1}{2}\,\mathbf{w}^T\mathbf{S}_0^{-1}\mathbf{w} \ + \ \mathbf{w}^T\mathbf{S}_0^{-1}\mathbf{m}_0 \ - \ \frac{1}{2}\,\mathbf{m}_0^T\mathbf{S}_0^{-1}\mathbf{m}_0 \\
&= -\frac{\beta}{2}\sum_{n=1}^{N}\mathbf{w}^T\phi(x_n)\,\phi(x_n)^T\mathbf{w} \ - \ \frac{1}{2}\,\mathbf{w}^T\mathbf{S}_0^{-1}\mathbf{w} \ + \ \beta\sum_{n=1}^{N} t_n\,(\mathbf{w}^T\phi(x_n)) \ + \ \mathbf{w}^T\mathbf{S}_0^{-1}\mathbf{m}_0 \ - \ \frac{\beta}{2}\sum_{n=1}^{N} t_n^2 \ - \ \frac{1}{2}\,\mathbf{m}_0^T\mathbf{S}_0^{-1}\mathbf{m}_0 \\
&= -\frac{1}{2}\,\mathbf{w}^T\Big(\sum_{n=1}^{N}\beta\phi(x_n)\,\phi(x_n)^T + \mathbf{S}_0^{-1}\Big)\mathbf{w} \ + \ \mathbf{w}^T\Big(\sum_{n=1}^{N}\beta\phi(x_n)\,t_n \ + \ \mathbf{S}_0^{-1}\mathbf{m}_0\Big) + \Big(- \frac{\beta}{2}\sum_{n=1}^{N} t_n^2 \ - \ \frac{1}{2}\,\mathbf{m}_0^T\mathbf{S}_0^{-1}\mathbf{m}_0\Big) \\
&= -\frac{1}{2}\,\mathbf{w}^T\Big(\beta\,\Phi^T\Phi + \mathbf{S}_0^{-1}\Big)\mathbf{w} \ + \ \mathbf{w}^T\Big(\beta\,\Phi^T\mathbf{t} + \mathbf{S}_0^{-1}\mathbf{m}_0\Big) + \Big(- \frac{\beta}{2}\sum_{n=1}^{N} t_n^2 \ - \ \frac{1}{2}\,\mathbf{m}_0^T\mathbf{S}_0^{-1}\mathbf{m}_0\Big) \longrightarrow (1)
\end{aligned}
$$

$$
\ln\mathcal{N}(\mathbf{w} \mid \mathbf{m}_N, \mathbf{S}_N) = -\frac{1}{2}\,\mathbf{w}^T\mathbf{S}_N^{-1}\mathbf{w} \ + \ \mathbf{w}^T\mathbf{S}_N^{-1}\mathbf{m}_N \ - \ \frac{1}{2}\,\mathbf{m}_N^T\mathbf{S}_N^{-1}\mathbf{m}_N \longrightarrow (2)
$$

From (1) and (2), comparing the quadratic and linear terms:

$$
\begin{aligned}
-\frac{1}{2}\,\mathbf{w}^T\mathbf{S}_N^{-1}\mathbf{w} &= -\frac{1}{2}\,\mathbf{w}^T\Big(\beta\,\Phi^T\Phi + \mathbf{S}_0^{-1}\Big)\mathbf{w}\\
&\Rightarrow \mathbf{S}_N^{-1} = \beta\,\Phi^T\Phi + \mathbf{S}_0^{-1}
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{w}^T\mathbf{S}_N^{-1}\mathbf{m}_N &= \mathbf{w}^T\Big(\beta\,\Phi^T\mathbf{t} + \mathbf{S}_0^{-1}\mathbf{m}_0\Big)\\
\Rightarrow \mathbf{S}_N^{-1}\mathbf{m}_N &= \Big(\beta\,\Phi^T\mathbf{t} + \mathbf{S}_0^{-1}\mathbf{m}_0\Big)\\
\Rightarrow \mathbf{m}_N &= \mathbf{S}_N\Big(\beta\,\Phi^T\mathbf{t} + \mathbf{S}_0^{-1}\mathbf{m}_0\Big)
\end{aligned}
$$

3.

$$
\begin{aligned}
p(\mathbf{w},\beta) &\;=\; \mathcal{N}(\mathbf{w}\mid\mathbf{m}_0,\beta^{-1}\mathbf{S}_0)\Gamma(\beta\mid a_0,b_0)\\
p(\mathbf{w},\beta\mid\mathbf{t}) &\;=\; \mathcal{N}(\mathbf{w}\mid\mathbf{m}_N,\beta^{-1}\mathbf{S}_N)\Gamma(\beta\mid a_N,b_N)\\
p(\mathbf{w},\beta\mid\mathbf{t}) &\;\propto\; p(\mathbf{t}\mid\mathbf{w},\beta)\,p(\mathbf{w},\beta)\\
p(\mathbf{t}\mid\mathbf{w},\beta) &= \prod_{n=1}^{N}\mathcal{N}\big(t_n\mid\mathbf{w}^T\phi(x_n),\beta^{-1}\big)
\end{aligned}
$$

$$
\mathcal{N}(\mathbf{x}\mid\boldsymbol{\mu},\Sigma) \;=\; \frac{1}{(2\pi)^{D/2}\,|\Sigma|^{1/2}}\,\exp\Big\{-\frac{1}{2}\,(\mathbf{x}-\boldsymbol{\mu})^\top\,\Sigma^{-1}\,(\mathbf{x}-\boldsymbol{\mu})\Big\}.
$$

$$
\begin{aligned}
\ln\mathcal{N}(\mathbf{w}\mid\mathbf{m}_0,\beta^{-1}\mathbf{S}_0) &= \ln\frac{1}{(2\pi)^{D/2}\,|\beta^{-1}\mathbf{S}_0|^{1/2}}\,\exp\Big\{-\frac{1}{2}\,(\mathbf{w}-\mathbf{m}_0)^T\,\beta^{-1}\mathbf{S}_0{}^{-1}\,(\mathbf{w}-\mathbf{m}_0)\Big\}\\
&= \ln\frac{1}{(2\pi)^{D/2}\,|\beta^{-1}\mathbf{S}_0|^{1/2}}\,\exp\Big\{-\frac{1}{2}\,(\mathbf{w}-\mathbf{m}_0)^T\,\beta\mathbf{S}_0{}^{-1}\,(\mathbf{w}-\mathbf{m}_0)\Big\}\\
&= \ln\frac{1}{(2\pi)^{D/2}\,|\beta^{-1}\mathbf{S}_0|^{1/2}} + \ln\exp\Big\{-\frac{1}{2}\,(\mathbf{w}-\mathbf{m}_0)^T\,\beta\mathbf{S}_0{}^{-1}\,(\mathbf{w}-\mathbf{m}_0)\Big\}\\
&= -\frac{D}{2}\,\ln(2\pi) - \frac{1}{2}\,\ln|\beta\mathbf{S}_0| - \frac{1}{2}(\mathbf{w}-\mathbf{m}_0)^T\beta\mathbf{S}_0^{-1}(\mathbf{w}-\mathbf{m}_0)\\
&= -\frac{1}{2}(\mathbf{w}-\mathbf{m}_0)^T\beta\mathbf{S}_0^{-1}(\mathbf{w}-\mathbf{m}_0)\\
&= -\frac{1}{2}\,\mathbf{w}^T\beta\mathbf{S}_0^{-1}\mathbf{w} \;+\; \mathbf{w}^T\beta\mathbf{S}_0^{-1}\mathbf{m}_0 \;-\; \frac{1}{2}\,\mathbf{m}_0^T\beta\mathbf{S}_0^{-1}\mathbf{m}_0
\end{aligned}
$$

$$
\begin{aligned}
\ln\Gamma(\beta\mid a_0,b_0) &= \ln\left(\frac{b_0^{a_0}}{\Gamma(a_0)}\beta^{a_0-1}\exp(-b_0\,\beta)\right)\\
&= \ln\frac{b_0^{a_0}}{\Gamma(a_0)} + \ln\beta^{a_0-1} + \ln\exp(-b_0\,\beta)\\
&= \ln b_0^{a_0} - \ln\Gamma(a_0) + (a_0-1)\ln\beta - b_0\beta\\
&= a_0\ln b_0 - \ln\Gamma(a_0) + (a_0-1)\ln\beta - b_0\beta
\end{aligned}
$$

$$
\begin{aligned}
\ln \prod_{n=1}^{N} \mathcal{N}\big(t_n \mid \mathbf{w}^T \phi(x_n),\, \beta^{-1}\big) &= \sum_{n=1}^{N} \ln \mathcal{N}\big(t_n \mid \mathbf{w}^T \phi(x_n), \beta^{-1}\big) \\
&= \sum_{n=1}^{N} \ln \left(\frac{1}{(2\pi)^{1/2}(\beta^{-1})^{1/2}} \exp\Big\{ -\frac{1}{2}(t_n - \mathbf{w}^T \phi(x_n))^2 \beta \Big\} \right) \\
&= \sum_{n=1}^{N} \left[\ln \frac{1}{(2\pi)^{1/2}(\beta^{-1})^{1/2}} + \ln \exp\Big\{ -\frac{1}{2}(t_n - \mathbf{w}^T \phi(x_n))^2 \beta \Big\} \right] \\
&= \sum_{n=1}^{N} \left[-\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(\beta^{-1}) - \frac{1}{2}(t_n - \mathbf{w}^T \phi(x_n))^2 \beta \right] \\
&= -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln(\beta^{-1}) - \frac{1}{2}\beta \sum_{n=1}^{N} (t_n - \mathbf{w}^T \phi(x_n))^2 \\
&= -\frac{N}{2} \ln(2\pi) + \frac{N}{2} \ln \beta - \frac{1}{2}\beta \sum_{n=1}^{N} (t_n - \mathbf{w}^T \phi(x_n))^2 \\
&= -\frac{\beta}{2} \sum_{n=1}^{N} \big(t_n \;-\; \mathbf{w}^T \phi(x_n) \big)^2 \\
&= -\frac{\beta}{2} \sum_{n=1}^{N} \big(t_n^2 \;-\; 2\, t_n \,(\mathbf{w}^T \phi(x_n)) \;+\; (\mathbf{w}^T \phi(x_n))^2 \big) \\
&= -\frac{\beta}{2} \sum_{n=1}^{N} \big(t_n^2 \;-\; 2\, t_n \,(\mathbf{w}^T \phi(x_n)) \;+\; \mathbf{w}^T \phi(x_n)\, \phi(x_n)^T \mathbf{w} \big) \\
&= -\frac{\beta}{2} \sum_{n=1}^{N} t_n^2 \;+\; \beta \sum_{n=1}^{N} t_n \left(\mathbf{w}^T \phi(x_n) \right) \;-\; \frac{\beta}{2} \sum_{n=1}^{N} \mathbf{w}^T \phi(x_n)\, \phi(x_n)^T \mathbf{w}
\end{aligned}
$$

$$
\begin{aligned}
\ln p(\mathbf{w}, \beta \mid \mathbf{t}) &= \ln\Big[p(\mathbf{t} \mid \mathbf{w}, \beta)\, p(\mathbf{w}, \beta)\Big] \\
&= \ln p(\mathbf{t} \mid \mathbf{w}, \beta) + \ln p(\mathbf{w}, \beta) \\
&= \ln \prod_{n=1}^{N} \mathcal{N}\Big(t_n \mid \mathbf{w}^T\phi(x_n),\ \beta^{-1}\Big) + \ln \mathcal{N}\Big(\mathbf{w} \mid \mathbf{m}_0,\ \beta^{-1}\mathbf{S}_0\Big) + \ln\Gamma\Big(\beta \mid a_0, b_0\Big) \\
&\propto\ -\frac{\beta}{2}\sum_{n=1}^{N}\Big[t_n^2 - 2\,t_n\,(\mathbf{w}^T\phi(x_n)) + \big(\mathbf{w}^T\phi(x_n)\big)^2\Big] \\
&\quad - \frac{\beta}{2}(\mathbf{w}-\mathbf{m}_0)^T\mathbf{S}_0^{-1}(\mathbf{w}-\mathbf{m}_0) + (a_0-1)\ln\beta - b_0\beta + a_0\ln b_0 - \ln\Gamma(a_0) \\
&= -\frac{\beta}{2}\sum_{n=1}^{N} t_n^2 + \beta\sum_{n=1}^{N} t_n\,(\mathbf{w}^T\phi(x_n)) - \frac{\beta}{2}\sum_{n=1}^{N}\Big(\mathbf{w}^T\phi(x_n)\Big)^2 \\
&\quad - \frac{\beta}{2}\,\mathbf{w}^T\mathbf{S}_0^{-1}\mathbf{w} + \beta\,\mathbf{w}^T\mathbf{S}_0^{-1}\mathbf{m}_0 - \frac{\beta}{2}\,\mathbf{m}_0^T\mathbf{S}_0^{-1}\mathbf{m}_0 \\
&\quad + (a_0-1)\ln\beta - b_0\beta + a_0\ln b_0 - \ln\Gamma(a_0) \\
&= -\frac{\beta}{2}\,\mathbf{w}^T\left(\sum_{n=1}^{N}\phi(x_n)\phi(x_n)^T + \mathbf{S}_0^{-1}\right)\mathbf{w} \\
&\quad + \beta\,\mathbf{w}^T\left(\sum_{n=1}^{N} t_n\,\phi(x_n) + \mathbf{S}_0^{-1}\mathbf{m}_0\right) - \frac{\beta}{2}\sum_{n=1}^{N} t_n^2 - \frac{\beta}{2}\,\mathbf{m}_0^T\mathbf{S}_0^{-1}\mathbf{m}_0 \\
&\quad + (a_0-1)\ln\beta - b_0\beta + a_0\ln b_0 - \ln\Gamma(a_0) \\
&= -\frac{\beta}{2}\,\mathbf{w}^T\Big(\Phi^T\Phi + \mathbf{S}_0^{-1}\Big)\mathbf{w} + \beta\,\mathbf{w}^T\Big(\Phi^T\mathbf{t} + \mathbf{S}_0^{-1}\mathbf{m}_0\Big) \\
&\quad - \frac{\beta}{2}\sum_{n=1}^{N} t_n^2 - \frac{\beta}{2}\,\mathbf{m}_0^T\mathbf{S}_0^{-1}\mathbf{m}_0 + (a_0-1)\ln\beta - b_0\beta + a_0\ln b_0 - \ln\Gamma(a_0) \longrightarrow (1)
\end{aligned}
$$

$$
\begin{aligned}
\ln p(\mathbf{w}, \beta \mid \mathbf{t}) &= \ln\Big[\mathcal{N}(\mathbf{w} \mid \mathbf{m}_N, \beta^{-1}\mathbf{S}_N)\,\Gamma(\beta \mid a_N, b_N)\Big] \\
&= \ln \mathcal{N}(\mathbf{w} \mid \mathbf{m}_N, \beta^{-1}\mathbf{S}_N)\ +\ \ln\Gamma(\beta \mid a_N, b_N) \\
&= -\frac{1}{2}(\mathbf{w}-\mathbf{m}_N)^T\beta\mathbf{S}_N^{-1}(\mathbf{w}-\mathbf{m}_N)\ +\ a_N\ln b_N - \ln\Gamma(a_N) + (a_N-1)\ln\beta - b_N\beta \\
&= -\frac{\beta}{2}\,\mathbf{w}^T\mathbf{S}_N^{-1}\mathbf{w}\ +\ \beta\,\mathbf{w}^T\mathbf{S}_N^{-1}\mathbf{m}_N \\
&\quad -\ \frac{\beta}{2}\,\mathbf{m}_N^T\mathbf{S}_N^{-1}\mathbf{m}_N\ +\ a_N\ln b_N - \ln\Gamma(a_N) +\ (a_N-1)\ln\beta - b_N\beta \longrightarrow (2)
\end{aligned}
$$

From (1) and (2), comparing the quadratic, linear and constant terms:

$$
\begin{aligned}
-\frac{\beta}{2}\,\mathbf{w}^T\mathbf{S}_N^{-1}\mathbf{w} &= -\frac{\beta}{2}\,\mathbf{w}^T\Big(\Phi^T\Phi + \mathbf{S}_0^{-1}\Big)\mathbf{w} \\
\Rightarrow \mathbf{S}_N^{-1} &= \Big(\Phi^T\Phi + \mathbf{S}_0^{-1}\Big) \\
\Rightarrow \mathbf{S}_N &= \Big(\Phi^T\Phi + \mathbf{S}_0^{-1}\Big)^{-1}
\end{aligned}
$$

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

$$
\begin{aligned}
-\frac{\beta}{2}\,\mathbf{m}_N^T\mathbf{S}_N^{-1}\mathbf{m}_N - b_N\beta &= -\frac{\beta}{2}\sum_{n=1}^{N} t_n^2 - \frac{\beta}{2}\,\mathbf{m}_0^T\mathbf{S}_0^{-1}\mathbf{m}_0 - b_0\beta\\
\Rightarrow -b_N\beta &= -\frac{\beta}{2}\sum_{n=1}^{N} t_n^2 - \frac{\beta}{2}\,\mathbf{m}_0^T\mathbf{S}_0^{-1}\mathbf{m}_0 - b_0\beta \;+\; \frac{\beta}{2}\,\mathbf{m}_N^T\mathbf{S}_N^{-1}\mathbf{m}_N\\
\Rightarrow b_N\beta &= \frac{\beta}{2}\sum_{n=1}^{N} t_n^2 + \frac{\beta}{2}\,\mathbf{m}_0^T\mathbf{S}_0^{-1}\mathbf{m}_0 + b_0\beta \;-\; \frac{\beta}{2}\,\mathbf{m}_N^T\mathbf{S}_N^{-1}\mathbf{m}_N\\
\Rightarrow b_N &= \frac{1}{2}\sum_{n=1}^{N} t_n^2 + \frac{1}{2}\,\mathbf{m}_0^T\mathbf{S}_0^{-1}\mathbf{m}_0 + b_0 \;-\; \frac{1}{2}\,\mathbf{m}_N^T\mathbf{S}_N^{-1}\mathbf{m}_N\\
\Rightarrow b_N &= b_0 + \frac{1}{2}\Big(\sum_{n=1}^{N} t_n^2 + \mathbf{m}_0^T\mathbf{S}_0^{-1}\mathbf{m}_0 - \mathbf{m}_N^T\mathbf{S}_N^{-1}\mathbf{m}_N\Big)
\end{aligned}
$$

$$
\begin{aligned}
p(\mathbf{t}\mid\mathbf{w},\beta) &= \prod_{n=1}^{N}\mathcal{N}\big(t_n\mid\mathbf{w}^T\phi(x_n),\beta^{-1}\big)\\
\Rightarrow p(\mathbf{t}\mid\mathbf{w},\beta) &= \prod_{n=1}^{N}\sqrt{\frac{\beta}{2\pi}}\exp\left(-\frac{\beta}{2}(t_n-\mathbf{w}^T\phi(x_n))^2\right)\\
\Rightarrow p(\mathbf{t}\mid\mathbf{w},\beta) &= \beta^{\frac{N}{2}}\cdot(2\pi)^{-\frac{N}{2}}\exp\left(-\frac{\beta}{2}\sum_{n=1}^{N}(t_n-\mathbf{w}^T\phi(x_n))^2\right)\\
\Rightarrow p(\mathbf{t}\mid\mathbf{w},\beta) &= \beta^{\frac{N}{2}}\exp\left(-\frac{\beta}{2}\|\mathbf{t}-\Phi\mathbf{w}\|^2\right)
\end{aligned}
$$

$$
\begin{aligned}
p(\mathbf{t}\mid\mathbf{w},\beta) &\propto \beta^{\frac{N}{2}}\exp\left(-\frac{\beta}{2}\|\mathbf{t}-\Phi\mathbf{w}\|^2\right)\to(3)\\
\Gamma(\beta\mid a_0,b_0) &\propto \beta^{a_0-1}\exp(-b_0\,\beta)\to(4)
\end{aligned}
$$

From (3) and (4)

$$
\beta^{a_0-1}\times\beta^{\frac{N}{2}} = \beta^{a_0-1+\frac{N}{2}}
$$

For the case where the likelihood consists of N univariate Gaussian terms where each term contributes a factor of $\beta^{\frac{1}{2}}$ when multiplying with N results in factor $\beta^{\frac{N}{2}}$, which increases the parameter a_0 by $\frac{N}{2}$. Therefore,

$$
\begin{aligned}
(a_N-1)\ln\beta &= \left(a_0-1+\frac{N}{2}\right)\ln\beta\\
\Rightarrow\quad a_N &= a_0+\frac{N}{2}
\end{aligned}
$$