9.21

$$(9.62)\quad \mathbb{E}[\ln p(\mathbf{t},\mathbf{w}|\alpha,\beta)] = \frac{M}{2}\ln\frac{\alpha}{2\pi} - \frac{\alpha}{2}\mathbb{E}[\mathbf{w}^T\mathbf{w}] + \frac{N}{2}\ln\frac{\beta}{2\pi} - \frac{\beta}{2}\sum_{n=1}^{N}\mathbb{E}[(t_n-\mathbf{w}^T\boldsymbol{\phi}_n)^2]$$

$$(3.49)\quad p(\mathbf{w}|\mathbf{t}) = \mathcal{N}(\mathbf{w}|\mathbf{m}_N, \mathbf{S}_N)$$

βについてEを最大化する。ただし$\mathbf{m}_N$, $\mathbf{S}_N$はα,βについて定数と考える。

(9.62)をβで微分して0とおく

$$\frac{\partial E}{\partial \beta} = \frac{N}{2}\frac{1}{\beta} - \frac{1}{2}\sum_n \mathbb{E}[(t_n-\mathbf{w}^T\boldsymbol{\phi}_n)^2] = 0 \quad \cdots ①$$

ここで

$$\begin{aligned}
\mathbb{E}[(t_n-\mathbf{w}^T\boldsymbol{\phi}_n)^2] &= \mathbb{E}[t_n^2 - 2t_n\mathbf{w}^T\boldsymbol{\phi}_n + \boldsymbol{\phi}_n^T\mathbf{w}\mathbf{w}^T\boldsymbol{\phi}_n] \\
&= \mathbb{E}[t_n]^2 - 2t_n\mathbb{E}[\mathbf{w}]^T\boldsymbol{\phi}_n + \boldsymbol{\phi}_n^T\mathbb{E}[\mathbf{w}\mathbf{w}^T]\boldsymbol{\phi}_n \\
&= t_n^2 - 2t_n\overset{(3.59)}{\mathbf{m}_N^T}\boldsymbol{\phi}_n + \boldsymbol{\phi}_n^T(\overset{(2.62)}{\mathbf{m}_N\mathbf{m}_N^T + \mathbf{S}_N})\boldsymbol{\phi}_n \\
&= (t_n - \mathbf{m}_N^T\boldsymbol{\phi}_n)^2 + \boldsymbol{\phi}_n^T\mathbf{S}_N\boldsymbol{\phi}_n
\end{aligned}$$

なので ① は

$$\frac{N}{\beta} - \sum_n\{(t_n-\mathbf{m}_N^T\boldsymbol{\phi}_n)^2 + \boldsymbol{\phi}_n^T\mathbf{S}_N\boldsymbol{\phi}_n\} = 0$$

となる。

これより

$$\begin{aligned}
\frac{1}{\beta} &= \frac{1}{N}\sum_n\{(t_n-\mathbf{m}_N^T\boldsymbol{\phi}_n)^2 + \boldsymbol{\phi}_n^T\mathbf{S}_N\boldsymbol{\phi}_n\} \\
&= \frac{1}{N}\left\{\sum_n(t_n-\mathbf{m}_N^T\boldsymbol{\phi}_n)^2 + \mathrm{Tr}(\boldsymbol{\Phi}\mathbf{S}_N\boldsymbol{\Phi}^T)\right\}
\end{aligned}$$

を得る。

(3.16)より

$$\boldsymbol{\Phi} = \underbrace{\begin{pmatrix}\boldsymbol{\phi}_1^T \\ \vdots \\ \boldsymbol{\phi}_N^T\end{pmatrix}}_{M次元}\Bigg\}N個$$

ブロック行列の積

$$\boldsymbol{\Phi}\mathbf{S}_N\boldsymbol{\Phi}^T = \begin{pmatrix}\boldsymbol{\phi}_1^T \\ \vdots \\ \boldsymbol{\phi}_N^T\end{pmatrix}\mathbf{S}_N(\boldsymbol{\phi}_1 \cdots \boldsymbol{\phi}_N) = \begin{pmatrix}\boldsymbol{\phi}_1^T\mathbf{S}_N \\ \vdots \\ \boldsymbol{\phi}_N^T\mathbf{S}_N\end{pmatrix}(\boldsymbol{\phi}_1 \cdots \boldsymbol{\phi}_N) = \begin{pmatrix}\boldsymbol{\phi}_1^T\mathbf{S}_N\boldsymbol{\phi}_1 & \cdots & \boldsymbol{\phi}_1^T\mathbf{S}_N\boldsymbol{\phi}_N \\ \vdots & \ddots & \vdots \\ \boldsymbol{\phi}_N^T\mathbf{S}_N\boldsymbol{\phi}_1 & \cdots & \boldsymbol{\phi}_N^T\mathbf{S}_N\boldsymbol{\phi}_N\end{pmatrix}$$

これより

$$\mathrm{Tr}(\boldsymbol{\Phi}\mathbf{S}_N\boldsymbol{\Phi}^T) = \boldsymbol{\phi}_1^T\mathbf{S}_N\boldsymbol{\phi}_1 + \cdots + \boldsymbol{\phi}_N^T\mathbf{S}_N\boldsymbol{\phi}_N = \sum_n \boldsymbol{\phi}_n^T\mathbf{S}_N\boldsymbol{\phi}_N$$