Chapter 3 Linear Models For Regression

Yue Yu

3.1 Linear Basis Function Models

(3.8)

Recall $t = y(\mathbf{x}, \mathbf{w}) + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \beta^{-1})$. This is equivalent to

$$\begin{aligned} p(\epsilon; \beta) &= \sqrt{\frac{\beta}{2\pi}} \exp\left\{ -\frac{\beta}{2}\epsilon^2 \right\} \\ &= \sqrt{\frac{\beta}{2\pi}} \exp\left\{ -\frac{\beta}{2}(t - y(\mathbf{x}, \mathbf{w}))^2 \right\}, \end{aligned}$$

which implies that

$$p(t|\mathbf{x}, \mathbf{w}, \beta) = \mathcal{N}(t|y(\mathbf{x}, \mathbf{w}), \beta^{-1}).$$

(3.13)

This equation should be

$$\nabla_{\mathbf{w}} \ln p(\mathbf{t}|\mathbf{w}, \beta) = \beta \sum_{n=1}^{N} \left\{ t_n - \mathbf{w}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}_n) \right\} \boldsymbol{\phi}(\mathbf{x}_n),$$

because

$$\nabla_{\mathbf{w}}(\mathbf{w}^{\mathrm{T}} \boldsymbol{\phi}(\mathbf{x}_n)) = \boldsymbol{\phi}(\mathbf{x}_n).$$

(3.14)

According to the fixed version of (3.13), this equation should be

$$\mathbf{0} = \sum_{n=1}^{N} t_n \boldsymbol{\phi}(\mathbf{x}_n) - \left(\sum_{n=1}^{N} \boldsymbol{\phi}(\mathbf{x}_n) \boldsymbol{\phi}(\mathbf{x}_n)^{\mathrm{T}} \right) \mathbf{w}.$$

(3.15)

By defining a design matrix $\boldsymbol{\Phi}$ in the form of (3.16), the fixed version of (3.14) can be reduced to

$$\mathbf{0} = \boldsymbol{\Phi}\mathbf{t} - \boldsymbol{\Phi}^{\mathrm{T}} \boldsymbol{\Phi} \mathbf{w}.$$

Solving for $\mathbf{w}$, we obtain

$$\mathbf{w}_{\mathrm{ML}} = (\boldsymbol{\Phi}^{\mathrm{T}} \boldsymbol{\Phi})^{-1} \boldsymbol{\Phi} \mathbf{t}.$$

(3.19)

$$\begin{aligned} \frac{\partial}{\partial w_0} E_D(\mathbf{w}) &= -\sum_{n=1}^{N} \left\{ t_n - w_0 - \sum_{j=1}^{M-1} w_j \phi_j(\mathbf{x}_n) \right\} \\ &= 0. \end{aligned}$$

Solving for w_0, we obtain

$$
\begin{aligned}
w_0 &= \frac{1}{N}\sum_{n=1}^{N} t_n - \frac{1}{N}\sum_{j=1}^{M-1} w_j \sum_{n=1}^{N}\phi_j(\mathbf{x}_n) \\
&= \bar{t} - \sum_{j=1}^{M-1} w_j \overline{\phi_j},
\end{aligned}
$$

where

$$
\bar{t} = \frac{1}{N}\sum_{n=1}^{N} t_n, \qquad \overline{\phi_j} = \frac{1}{N}\sum_{n=1}^{N}\phi_j(\mathbf{x}_n).
$$

(3.23)

$$
\begin{aligned}
\nabla_{\mathbf{w}} E_n(\mathbf{w}) &= -(t_n - \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n))\boldsymbol{\phi}(\mathbf{x}_n) \\
&= -(t_n - \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}_n)\boldsymbol{\phi}_n,
\end{aligned}
$$

where $\boldsymbol{\phi}_n = \boldsymbol{\phi}(\mathbf{x}_n)$. Plug it into (3.22), we obtain (3.23).

(3.28)

Recall the regularized error function

$$
\begin{aligned}
E(\mathbf{w}) &= \frac{1}{2}\sum_{n=1}^{N}\{t_n - \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n)\}^2 + \frac{\lambda}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} \\
&= \frac{1}{2}\big(\mathbf{t} - \boldsymbol{\Phi}\mathbf{w}\big)^{\mathrm{T}}\big(\mathbf{t} - \boldsymbol{\Phi}\mathbf{w}\big) + \frac{\lambda}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}.
\end{aligned}
$$

Setting the gradient with respect to $\mathbf{w}$ to $\mathbf{0}$, we have

$$
\begin{aligned}
\nabla_{\mathbf{w}} E(\mathbf{w}) &= \frac{1}{2}\nabla_{\mathbf{w}}\big(\mathbf{t}^{\mathrm{T}}\mathbf{t} - 2\mathbf{t}^{\mathrm{T}}\boldsymbol{\Phi}\mathbf{w} + \mathbf{w}^{\mathrm{T}}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}\mathbf{w} + \lambda\mathbf{w}^{\mathrm{T}}\mathbf{w}\big) \\
&= \big(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi} + \lambda\mathbf{I}\big)\mathbf{w} - \boldsymbol{\Phi}^{\mathrm{T}}\mathbf{t} \\
&= \mathbf{0}.
\end{aligned}
$$

Solving for $\mathbf{w}$, we obtain

$$
\mathbf{w} = \big(\lambda\mathbf{I} + \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}\big)^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{t}.
$$

(3.40)

$$
\begin{aligned}
&\mathbb{E}_{\mathcal{D}}\big[\{y(\mathbf{x};\mathcal{D}) - h(\mathbf{x})\}^2\big] \\
&= \mathbb{E}_{\mathcal{D}}\big[\{y(\mathbf{x};\mathcal{D}) - \mathbb{E}_{\mathcal{D}}[y(\mathbf{x};\mathcal{D})] + \mathbb{E}_{\mathcal{D}}[y(\mathbf{x};\mathcal{D})] - h(\mathbf{x})\}^2\big] \\
&= \mathbb{E}_{\mathcal{D}}\big[\{y(\mathbf{x};\mathcal{D}) - \mathbb{E}_{\mathcal{D}}[y(\mathbf{x};\mathcal{D})]\}^2 + \{\mathbb{E}_{\mathcal{D}}[y(\mathbf{x};\mathcal{D})] - h(\mathbf{x})\}^2 + 2\{y(\mathbf{x};\mathcal{D}) - \mathbb{E}_{\mathcal{D}}[y(\mathbf{x};\mathcal{D})]\}\{\mathbb{E}_{\mathcal{D}}[y(\mathbf{x};\mathcal{D})] - h(\mathbf{x})\}\big] \\
&= \mathbb{E}_{\mathcal{D}}\big[\{\mathbb{E}_{\mathcal{D}}[y(\mathbf{x};\mathcal{D})] - h(\mathbf{x})\}^2\big] + \mathbb{E}_{\mathcal{D}}\big[\{y(\mathbf{x};\mathcal{D}) - \mathbb{E}_{\mathcal{D}}[y(\mathbf{x};\mathcal{D})]\}^2\big] + 2\mathbb{E}_{\mathcal{D}}\big[y(\mathbf{x};\mathcal{D})\mathbb{E}_{\mathcal{D}}[y(\mathbf{x};\mathcal{D})] - y(\mathbf{x};\mathcal{D})h(\mathbf{x}) - \\
&\quad \mathbb{E}_{\mathcal{D}}[y(\mathbf{x};\mathcal{D})]^2 + \mathbb{E}_{\mathcal{D}}[y(\mathbf{x};\mathcal{D})]h(\mathbf{x})\big] \\
&= \{\mathbb{E}_{\mathcal{D}}[y(\mathbf{x};\mathcal{D})] - h(\mathbf{x})\}^2 + \mathbb{E}_{\mathcal{D}}\big[\{y(\mathbf{x};\mathcal{D}) - \mathbb{E}_{\mathcal{D}}[y(\mathbf{x};\mathcal{D})]\}^2\big] + 2\{\mathbb{E}_{\mathcal{D}}[y(\mathbf{x};\mathcal{D})]^2 - \mathbb{E}_{\mathcal{D}}[y(\mathbf{x};\mathcal{D})]h(\mathbf{x}) - \\
&\quad \mathbb{E}_{\mathcal{D}}[y(\mathbf{x};\mathcal{D})]^2 + \mathbb{E}_{\mathcal{D}}[y(\mathbf{x};\mathcal{D})]h(\mathbf{x})\} \\
&= \{\mathbb{E}_{\mathcal{D}}[y(\mathbf{x};\mathcal{D})] - h(\mathbf{x})\}^2 + \mathbb{E}_{\mathcal{D}}\big[\{y(\mathbf{x};\mathcal{D}) - \mathbb{E}_{\mathcal{D}}[y(\mathbf{x};\mathcal{D})]\}^2\big].
\end{aligned}
$$

(3.50) and (3.51)

$$\begin{aligned}\mathcal{N}(\mathbf{w}|\mathbf{m}_N, \mathbf{S}_N) &= p(\mathbf{w}|\mathbf{t}) \\ &\propto p(\mathbf{t}|\mathbf{w})p(\mathbf{w}) \\ &= \mathcal{N}(\mathbf{t}|\mathbf{X}, \mathbf{w}, \beta)\mathcal{N}(\mathbf{w}|\mathbf{m}_0, \mathbf{S}_0) \\ &= \left(\prod_{n=1}^{N} \mathcal{N}(t_n|\mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n), \beta^{-1})\right)\mathcal{N}(\mathbf{w}|\mathbf{m}_0, \mathbf{S}_0). \end{aligned} \tag{*}$$

Consider the exponential term

$$\begin{aligned}&-\frac{\beta}{2}\sum_{n=1}^{N}(t_n - \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n))^2 - \frac{1}{2}(\mathbf{w}-\mathbf{m}_0)^{\mathrm{T}}\mathbf{S}_0^{-1}(\mathbf{w}-\mathbf{m}_0) \\ =& -\frac{\beta}{2}(\mathbf{t}-\boldsymbol{\Phi}\mathbf{w})^{\mathrm{T}}(\mathbf{t}-\boldsymbol{\Phi}\mathbf{w}) - \frac{1}{2}(\mathbf{w}-\mathbf{m}_0)^{\mathrm{T}}\mathbf{S}_0^{-1}(\mathbf{w}-\mathbf{m}_0),\end{aligned}$$

by matching the coefficients of the quadratic terms on both sides of (*), we have

$$\mathbf{S}_N^{-1} = \mathbf{S}_0^{-1} + \beta\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}.$$

Similarly, matching the coefficients of the linear terms on both sides of (*), we have

$$\mathbf{m}_N^{\mathrm{T}}\mathbf{S}_N^{-1} = \beta\mathbf{t}^{\mathrm{T}}\boldsymbol{\Phi} + \mathbf{m}_0^{\mathrm{T}}\mathbf{S}_0^{-1},$$

solving for $\mathbf{m}_N$, we obtain

$$\mathbf{m}_N = \mathbf{S}_N\big(\mathbf{S}_0^{-1}\mathbf{m}_0 + \beta\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{t}\big).$$

(3.57)

To be consistent with (3.58), this equation should be

$$p(t|\mathbf{x}, \mathbf{t}, \alpha, \beta) = \int p(t|\mathbf{x}, \mathbf{w}, \beta)p(\mathbf{w}|\mathbf{t}, \alpha, \beta)\,\mathrm{d}\mathbf{w}.$$

(3.58) and (3.59)

Recall (2.113) and (2.114), mapping $p(t|\mathbf{x}, \mathbf{w}, \beta)$ to $p(\mathbf{y}|\mathbf{x})$, and $p(\mathbf{w}|\mathbf{t}, \alpha, \beta)$ to $p(\mathbf{x})$, we have

$$\begin{aligned}p(t|\mathbf{x}, \mathbf{w}, \beta) &= \mathcal{N}(t|\mathbf{w}^{\mathrm{T}}\mathbf{x}, \beta^{-1}) \\ &= \mathcal{N}(t|\mathbf{A}\mathbf{w} + \mathbf{b}, \mathbf{L}^{-1}) \\ p(\mathbf{w}|\mathbf{t}, \alpha, \beta) &= \mathcal{N}(\mathbf{w}|\mathbf{m}_N, \mathbf{S}_N) \\ &= \mathcal{N}(\mathbf{w}|\boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1}),\end{aligned}$$

which implies that

$$\begin{aligned}\mathbf{A} &= \boldsymbol{\phi}(\mathbf{x})^{\mathrm{T}} \\ \mathbf{b} &= \mathbf{0} \\ \mathbf{L} &= \beta \\ \boldsymbol{\mu} &= \mathbf{m}_N \\ \boldsymbol{\Lambda} &= \mathbf{S}_N^{-1}.\end{aligned}$$

Substituting them back into (2.115), we obtain (3.58) and (3.59).

(3.63)

$$
\begin{aligned}
\operatorname{cov}[y(\mathbf{x}), y(\mathbf{x}')] &= \operatorname{cov}[\boldsymbol{\phi}(\mathbf{x})^{\mathrm{T}}\mathbf{w}, \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}')] \\
&= \mathbb{E}[\boldsymbol{\phi}(\mathbf{x})^{\mathrm{T}}\mathbf{w}\mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}')] - \mathbb{E}[\boldsymbol{\phi}(\mathbf{x})^{\mathrm{T}}\mathbf{w}]\mathbb{E}[\mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}')] \\
&= \boldsymbol{\phi}(\mathbf{x})^{\mathrm{T}}\mathbb{E}[\mathbf{w}\mathbf{w}^{\mathrm{T}}]\boldsymbol{\phi}(\mathbf{x}') - \boldsymbol{\phi}(\mathbf{x})^{\mathrm{T}}\mathbf{m}_N\mathbf{m}_N^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}') \\
&= \boldsymbol{\phi}(\mathbf{x})^{\mathrm{T}}(\operatorname{cov}[\mathbf{w}] + \mathbb{E}[\mathbf{w}]\mathbb{E}[\mathbf{w}^{\mathrm{T}}])\boldsymbol{\phi}(\mathbf{x}') - \boldsymbol{\phi}(\mathbf{x})^{\mathrm{T}}\mathbf{m}_N\mathbf{m}_N^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}') \\
&= \boldsymbol{\phi}(\mathbf{x})^{\mathrm{T}}(\operatorname{cov}[\mathbf{w}] + \mathbf{m}_N\mathbf{m}_N^{\mathrm{T}})\boldsymbol{\phi}(\mathbf{x}') - \boldsymbol{\phi}(\mathbf{x})^{\mathrm{T}}\mathbf{m}_N\mathbf{m}_N^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}') \\
&= \boldsymbol{\phi}(\mathbf{x})^{\mathrm{T}}\operatorname{cov}[\mathbf{w}]\boldsymbol{\phi}(\mathbf{x}') \\
&= \boldsymbol{\phi}(\mathbf{x})^{\mathrm{T}}\mathbf{S}_N\boldsymbol{\phi}(\mathbf{x}') \\
&= \beta^{-1}k(\mathbf{x}, \mathbf{x}').
\end{aligned}
$$

(3.67)

$$
\begin{aligned}
p(t|\mathbf{x}, \mathcal{D}) &= \sum_{i=1}^{L} p(t, \mathcal{M}_i|\mathbf{x}, \mathcal{D}) \\
&= \sum_{i=1}^{L} p(t|\mathbf{x}, \mathcal{M}_i, \mathcal{D})p(\mathcal{M}_i|\mathcal{D}).
\end{aligned}
$$

(3.72)

$$
\begin{aligned}
\ln p(\mathcal{D}) &\simeq \ln\left\{p(\mathcal{D}|\mathbf{w}_{\mathrm{MAP}})\left(\frac{\Delta w_{\mathrm{posterior}}}{\Delta w_{\mathrm{prior}}}\right)^{M}\right\} \\
&= \ln p(\mathcal{D}|\mathbf{w}_{\mathrm{MAP}}) + M\ln\left(\frac{\Delta w_{\mathrm{posterior}}}{\Delta w_{\mathrm{prior}}}\right).
\end{aligned}
$$

(3.78) and (3.79)

$$
\begin{aligned}
p(\mathbf{t}|\alpha, \beta) &= \int p(\mathbf{t}|\mathbf{w}, \beta)p(\mathbf{w}|\alpha)\,\mathrm{d}\mathbf{w} \\
&= \int \left[\prod_{n=1}^{N} \mathcal{N}(t_n|\mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n), \beta^{-1})\right]\mathcal{N}(\mathbf{w}|\mathbf{0}, \alpha^{-1}\mathbf{I})\,\mathrm{d}\mathbf{w} \\
&= \int \left[\prod_{n=1}^{N} \sqrt{\frac{\beta}{2\pi}}\exp\left\{-\frac{\beta}{2}(t_n - \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n))^2\right\}\right]\frac{1}{(2\pi)^{M/2}|\alpha^{-1}\mathbf{I}|^{1/2}}\exp\left\{-\frac{1}{2}\mathbf{w}^{\mathrm{T}}(\alpha^{-1}\mathbf{I})^{-1}\mathbf{w}\right\}\mathrm{d}\mathbf{w} \\
&= \left(\frac{\beta}{2\pi}\right)^{N/2}\left(\frac{\alpha}{2\pi}\right)^{M/2}\int \exp\left\{-\frac{\beta}{2}\sum_{n=1}^{N}(t_n - \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n))^2 - \frac{\alpha}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}\right\}\mathrm{d}\mathbf{w} \\
&= \left(\frac{\beta}{2\pi}\right)^{N/2}\left(\frac{\alpha}{2\pi}\right)^{M/2}\int \exp\left\{-\frac{\beta}{2}(\mathbf{t} - \boldsymbol{\Phi}\mathbf{w})^2 - \frac{\alpha}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}\right\}\mathrm{d}\mathbf{w} \\
&= \left(\frac{\beta}{2\pi}\right)^{N/2}\left(\frac{\alpha}{2\pi}\right)^{M/2}\int \exp\left\{-E(\mathbf{w})\right\}\mathrm{d}\mathbf{w},
\end{aligned}
$$

where $E(\mathbf{w})$ satisfies (3.79).

(3.80) to (3.84)

According to (3.79),

$$\begin{aligned}
E(\mathbf{w}) &= \beta E_D(\mathbf{w}) + \alpha E_W(\mathbf{w}) \\
&= \frac{\beta}{2}\|\mathbf{t} - \mathbf{\Phi}\mathbf{w}\|^2 + \frac{\alpha}{2}\mathbf{w}^{\mathbf{T}}\mathbf{w} \\
&= \frac{1}{2}(\mathbf{w}^{\mathrm{T}}(\alpha\mathbf{I} + \beta\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})\mathbf{w} - 2\beta\mathbf{t}^{\mathrm{T}}\mathbf{\Phi}\mathbf{w} + \beta\mathbf{t}^{\mathrm{T}}\mathbf{t}). \qquad (*)
\end{aligned}$$

Comparing with the quadratic term in (3.80), we have

$$\frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{A}\mathbf{w} = \mathbf{w}^{\mathrm{T}}(\alpha\mathbf{I} + \beta\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})\mathbf{w},$$

which implies that

$$\mathbf{A} = \alpha\mathbf{I} + \beta\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}. \qquad (**)$$

Then, by comparing with the linear term in (3.80), we have

$$\mathbf{m}_N^{\mathrm{T}}\mathbf{A} = \beta\mathbf{t}^{\mathrm{T}}\mathbf{\Phi}.$$

Noticing that $\mathbf{A}$ is symmetric, solving for $\mathbf{m}_N$, we obtain

$$\mathbf{m}_N = \beta\mathbf{A}^{-1}\mathbf{\Phi}^{\mathrm{T}}\mathbf{t}. \qquad (***)$$

Substitute (**) and (***) into (*), we have

$$\begin{aligned}
E(\mathbf{w}) &= \frac{1}{2}(\mathbf{w}^{\mathrm{T}}(\alpha\mathbf{I} + \beta\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})\mathbf{w} - 2\beta\mathbf{t}^{\mathrm{T}}\mathbf{\Phi}\mathbf{w} + \beta\mathbf{t}^{\mathrm{T}}\mathbf{t}) \\
&= \frac{1}{2}(\mathbf{w}^{\mathrm{T}}\mathbf{A}\mathbf{w} - 2\beta\mathbf{t}^{\mathrm{T}}\mathbf{\Phi}\mathbf{w} + \beta\mathbf{t}^{\mathrm{T}}\mathbf{t}) \\
&= \frac{1}{2}\{(\mathbf{w} - \mathbf{m}_N)^{\mathrm{T}}\mathbf{A}(\mathbf{w} - \mathbf{m}_N) + \mathbf{m}_N^{\mathrm{T}}\mathbf{A}\mathbf{w} - \mathbf{m}_N^{\mathrm{T}}\mathbf{A}\mathbf{m}_N - 2\beta\mathbf{t}^{\mathrm{T}}\mathbf{\Phi}\mathbf{w} + \beta\mathbf{t}^{\mathrm{T}}\mathbf{t}\} \\
&= \frac{1}{2}(\mathbf{w} - \mathbf{m}_N)^{\mathrm{T}}\mathbf{A}(\mathbf{w} - \mathbf{m}_N) + \frac{1}{2}(\beta\mathbf{t}^{\mathrm{T}}\mathbf{t} - \mathbf{m}_N^{\mathrm{T}}\mathbf{A}\mathbf{m}_N) \\
&= \frac{1}{2}(\mathbf{w} - \mathbf{m}_N)^{\mathrm{T}}\mathbf{A}(\mathbf{w} - \mathbf{m}_N) + \frac{1}{2}(\beta\mathbf{t}^{\mathrm{T}}\mathbf{t} - 2\mathbf{m}_N^{\mathrm{T}}\mathbf{A}\mathbf{m}_N + \mathbf{m}_N^{\mathrm{T}}\mathbf{A}\mathbf{m}_N) \\
&= \frac{1}{2}(\mathbf{w} - \mathbf{m}_N)^{\mathrm{T}}\mathbf{A}(\mathbf{w} - \mathbf{m}_N) + \frac{1}{2}\{\beta\mathbf{t}^{\mathrm{T}}\mathbf{t} - 2\mathbf{m}_N^{\mathrm{T}}\mathbf{A}(\beta\mathbf{A}^{-1}\mathbf{\Phi}^{\mathrm{T}}\mathbf{t}) + \mathbf{m}_N^{\mathrm{T}}(\alpha\mathbf{I} + \beta\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})\mathbf{m}_N\} \\
&= \frac{1}{2}(\mathbf{w} - \mathbf{m}_N)^{\mathrm{T}}\mathbf{A}(\mathbf{w} - \mathbf{m}_N) + \frac{1}{2}\{\beta\mathbf{t}^{\mathrm{T}}\mathbf{t} - 2\beta(\mathbf{\Phi}\mathbf{m}_N)^{\mathrm{T}}\mathbf{t} + \beta\mathbf{m}_N^{\mathrm{T}}\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}\mathbf{m}_N\} + \frac{\alpha}{2}\mathbf{m}_N^{\mathrm{T}}\mathbf{m}_N \\
&= \frac{1}{2}(\mathbf{w} - \mathbf{m}_N)^{\mathrm{T}}\mathbf{A}(\mathbf{w} - \mathbf{m}_N) + \frac{\beta}{2}\|\mathbf{t} - \mathbf{\Phi}\mathbf{m}_N\|^2 + \frac{\alpha}{2}\mathbf{m}_N^{\mathrm{T}}\mathbf{m}_N \\
&= E(\mathbf{m}_N) + \frac{1}{2}(\mathbf{w} - \mathbf{m}_N)^{\mathrm{T}}\mathbf{A}(\mathbf{w} - \mathbf{m}_N),
\end{aligned}$$

where $E(\mathbf{m}_N)$ satisfies (3.82).