Some Results from Section 3.1.1 and 3.3

- Standard Results

$$p(\underline{x}) = \mathcal{N}(\underline{x} \mid \underline{\mu}, \underline{\Lambda}^{-1}) \quad \cdots \quad (2.113)$$

$$p(\underline{y}|\underline{x}) = \mathcal{N}(\underline{y} \mid \underline{A}\underline{x}+\underline{b}, \underline{L}^{-1}) \quad \cdots \quad (2.114)$$

$$p(\underline{y}) = \mathcal{N}(\underline{y} \mid \underline{A}\underline{\mu}+\underline{b}, L^{-1}+A\Lambda^{-1}A^{T}) \cdots (2.115)$$

$$p(x|y) = \mathcal{N}(x \mid \Sigma\{A^{T}L(y-b)+\Lambda\mu\}, \Sigma) \cdots (2.116)$$

where $\Sigma = (\Lambda + A^{T}LA)^{-1}$

- Results from Section 3.1.1 and 3.3

$$t = y(\underline{x}, \underline{\omega}) + \epsilon \qquad \text{where } y(\underline{x},\underline{\omega}) = \underline{\omega}^{T}\underline{\phi}(\underline{x}) \qquad (3.7)$$

$$p(t \mid x, \omega, \beta) = \mathcal{N}(t \mid y(x,\omega), \beta^{-1}) \qquad (3.8)$$

$$X = \{x_1, x_2, \cdots x_N\}$$

Likelihood:

$$p(\underline{t} \mid \underline{X}, \underline{\omega}, \beta) = \prod_{n=1}^{N} \mathcal{N}(t_n \mid \omega^{T}\phi(x_n), \beta^{-1}) \cdots (3.10)$$

$$\underset{N\times M}{\Phi} = \begin{bmatrix} -\underline{\phi}^{T}(x_1)- \\ -\underline{\phi}^{T}(x_2)- \\ \vdots \\ -\underline{\phi}^{T}(x_N)- \end{bmatrix} \qquad \cdots (3.16)$$

Assumption: β is a constant and X is also fixed.

prior $\rightarrow$ $$p(\underline{\omega}) = \mathcal{N}(\underline{\omega} \mid m_0, S_0) \quad \cdots \quad (3.48)$$

$$p(\underline{t}|\underline{\omega}) = \prod_{n=1}^{N} \mathcal{N}(t_n \mid \underline{\phi}^{T}(x_n)\,\underline{\omega}, \beta^{-1})$$

$$\underline{t} = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{bmatrix} \qquad = \mathcal{N}(\underline{t} \mid \Phi\omega, \beta^{-1}I) \quad \cdots \quad (i)$$

Compare with (2.113) and (2.114), $\mu = m_0$, $\Lambda^{-1} = S_0$, $A = \Phi$, $x = \omega$, $b = 0$, $L^{-1} = \beta^{-1}I$

$$\therefore \; p(\omega|\underline{t}) = \mathcal{N}(\underline{\omega} \mid \Sigma\{A^{T}L(\underline{t}-\underline{b})+\Lambda\mu\}, \Sigma) \quad \text{where } \Sigma = (\Lambda + A^{T}LA)^{-1}$$

using 2.116

$$= \mathcal{N}(\underline{\omega} \mid \underline{m}_N, S_N) \quad \cdots \quad (3.49)$$

$$S_N = \Sigma = (S_0^{-1} + \beta\phi^{T}\phi)^{-1} \quad \cdots \quad (3.50)$$

$$m_N = S_N\{\beta\phi^{T}\underline{t} + S_0^{-1}m_0\} \quad \cdots \quad (3.51)$$

~~predictive distribution~~

$\cancel{p(t|\underline{t}, \alpha, \beta)}$

Let we have an isotropic prior

$$p(\underline{\omega}|\alpha) = \mathcal{N}(\underline{\omega}|\underline{0}, \alpha^{-1}I) \quad \cdots (3.52)$$

$$\therefore \quad \underline{m}_N = \beta S_N \underline{\Phi}^T \underline{t} \quad \cdots (3.53)$$

$$S_N^{-1} = \alpha I + \beta \underline{\Phi}^T \underline{\Phi} \quad \cdots (3.54)$$

predictive distribution

$$p(t|\underline{t}, \alpha, \beta) = \int_{\underline{\omega}} \underbrace{p(t|\underline{\omega}, \beta)}_{p(y|x)} \underbrace{p(\underline{\omega}|\underline{t}, \alpha, \beta)}_{p(x)} d\underline{\omega} \quad \cdots (3.57)$$

$$= \int_{\underline{\omega}} p(t, \underline{\omega}|\underline{t}, \alpha, \beta)\, d\underline{\omega}$$

$$\left.\begin{aligned} p(t|\underline{\omega}, \beta) &= \mathcal{N}(t|\phi^T(x)\omega, \beta^{-1}) \\ p(\underline{\omega}|\underline{t}, \alpha, \beta) &= \mathcal{N}(\underline{\omega}|\underline{m}_N, S_N) \end{aligned}\right\} \rightarrow \begin{aligned} A = \phi^T(x), \quad L^{-1} = \beta^{-1} \\ \underline{\mu} = \underline{m}_N, \quad \Lambda^{-1} = S_N \end{aligned}$$

$$\therefore \quad p(t|\underline{t}, \alpha, \beta) = \mathcal{N}(\omega|\phi^T(x) m_N, \sigma_N^2(x)) \quad \cdots (3.58)$$

use (2.115)

$$\text{where } \sigma_N^2(x) = \beta^{-1} + \phi^T(x) S_N \phi(x) \quad \cdots (3.59)$$

$$\underbrace{p(\underline{t} \mid \underline{X}, \underline{w}, \beta)}_{p(y|x)} = \prod_{n=1}^{N} p(t_n \mid \underline{x}_n, \underline{w}, \beta^{-1})$$

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

$$= \mathcal{N}(\underline{t} \mid \Phi w, \beta^{-1} I) \quad \cdots (7.79)$$

$$\underbrace{p(\underline{w} \mid \underline{\alpha})}_{p(x)} = \prod_{i=1}^{M} \mathcal{N}(w_i \mid 0, \alpha_i^{-1})$$

$$= \mathcal{N}(\underline{w} \mid \underline{0}, \mathrm{diag}\{\alpha_i^{-1}\}) \quad \cdots (7.80)$$

compare with (2.114) and (2.113), $A' = \Phi$, $L^{-1} = \beta^{-1} I$; $b = 0$; $\mu = 0$, $\Lambda^{-1} = \mathrm{diag}\{\alpha_i^{-1}\}$

$$\underbrace{p(\underline{w} \mid \underline{t}, \underline{X}, \underline{\alpha}, \beta)}_{p(x|y)} = \mathcal{N}(w \mid \underline{m}, \Sigma) \quad \cdots (7.81)$$

where $\Sigma = (\Lambda + A'^T L A')^{-1}$

$$= (\mathrm{diag}\{\alpha_i\} + \beta \Phi^T \Phi)^{-1}$$

$$= (A + \beta \Phi^T \Phi)^{-1} \quad \cdots (7.83)$$

here $A = \mathrm{diag}\{\alpha_i\}$; $\Lambda = A$, $\Lambda^{-1} = A^{-1}$

and $\underline{m} = \Sigma \{ A'^T L (\underline{t} - b) + \Lambda \underline{\mu} \}$

$$= \Sigma \{ \Phi^T \beta I \underline{t} \}$$

$$= \beta \Sigma \Phi^T \underline{t} \quad \cdots (7.82)$$

<u>evidence approximation</u>

$$p(\underline{t} \mid \underline{X}, \underline{\alpha}, \beta) = \int_{\underline{w}} \underbrace{p(\underline{t} \mid \underline{X}, \underline{w}, \beta)}_{p(y|x)} \underbrace{p(\underline{w} \mid \underline{\alpha})}_{p(x)} \, d\underline{w}$$

$$\underset{\text{use (2.115)}}{=} \mathcal{N}(\underline{t} \mid \underline{A'\mu} + \underline{b}, L^{-1} + A' \Lambda^{-1} A'^T)$$

$$= \mathcal{N}(\underline{t} \mid \underline{0}, \beta^{-1} I + \Phi A^{-1} \Phi^T) \quad \cdots (7.86)$$

$$= \mathcal{N}(\underline{t} \mid \underline{0}, \underline{C})$$

$$= \frac{1}{(2\pi)^{N/2}} \cdot \frac{1}{|C|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} \underline{t}^T C^{-1} \underline{t} \right\}$$

$$\therefore \ln p(\underline{t} \mid \underline{X}, \underline{\alpha}, \beta) = -\frac{N}{2} \ln(2\pi) - \frac{1}{2} \ln |C| - \frac{1}{2} \underline{t}^T C^{-1} \underline{t} \quad \cdots (7.85)$$

<u>predictive distribution</u>

$$p(t \mid \underline{x}, \underline{X}, \underline{t}, \alpha^*, \beta^*) = \int_{\underline{w}} \underbrace{p(t \mid \underline{x}, \underline{w}, \beta^*)}_{p(y|x)} \; \underbrace{p(\underline{w} \mid \underline{X}, \underline{t}, \alpha^*, \beta^*)}_{p(x)} \, d\underline{w}$$

$$\mathcal{N}(\underline{t} \mid \phi^T(x)\underline{w}, \beta^{*-1}) \qquad \mathcal{N}(\underline{w} \mid \underline{m}, \underline{\Sigma})$$

$$(7.76) \qquad\qquad (7.81)$$

Compare with (2.114) and (2.113), $A' = \phi^T(x)$ $b = 0$, $L^{-1} = \beta^{*-1}$; $\underline{\mu} = m$, $\Lambda^{-1} = \Sigma$.

$$\therefore \underbrace{p(t \mid \underline{x}, \underline{X}, \underline{t}, \alpha^*, \beta^*)}_{p(y)} \underset{(2.115)}{\overset{\uparrow}{=}} \mathcal{N}(t \mid A'\mu + b, \; L^{-1} + A'\Lambda^{-1}A'^T)$$

$$= \mathcal{N}(t \mid \phi^T(x)\, m, \; \beta^{*-1} + \phi^T(x)\, \Sigma\, \phi(x))$$

$$= \mathcal{N}(t \mid \underline{m}^T \underline{\phi}(\underline{x}), \; \sigma^2(\underline{x})) \quad \ldots \quad (7.90)$$

where $\sigma^2(\underline{x}) = \beta^{*-1} + \phi^T(x)\, \Sigma\, \phi(x)$ $\quad \ldots \quad$ (7.91)