3.7.

$\mathbf{t}$: target variable.

β : noise precison parameter. Constant.

$p(w) = \mathcal{N}(w \mid m_0, S_0)$; prior

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

이때 posterior 는

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

$\phi(x_n)^T$:

↓

scalar.

||

$\phi(x_n)^T w$

$$\propto (2\pi)^{\frac{K}{2}} |S_0|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(w-m_0)^T S_0^{-1}(w-m_0)\right)$$

$$\cdot (2\pi)^{\frac{NK}{2}} \cdot |\beta|^{\frac{N}{2}} \exp\left[\sum \frac{1}{2}(t_n - w^T\phi(x_n))^T \beta (t_n - w^T\phi(x_n))\right]$$

$$\propto \exp\left(\sum (t_n - w^T\phi(x_n))^T \beta (t_n - w^T\phi(x_n)) + (w-m_0)^T S_0^{-1}(w-m_0)\right)$$

exp 안을 전개하면

$$\sum (t_n - \phi(x_n)^T w)^T \beta (t_n - \phi(x_n)^T w) + (w-m_0)^T S_0^{-1}(w-m_0)$$

$$= \sum w^T\phi(x_n)\beta\phi(x_n)^T w - 2\beta\phi(x_n)^T w + w^T S_0^{-1} w - 2w^T S_0^{-1} m_0 + \square$$

$$= \sum (w^T\phi(x_n))\beta(\phi(x_n)^T w) + w^T S_0^{-1} w - 2t_n w^T\phi(x_n)\beta - 2w^T S_0^{-1} m_0 + \square$$

$$= w^T\left[\sum \phi(x_n)\beta\phi(x_n)^T + S_0^{-1}\right] w - \sum 2 w^T(\beta\phi(x_n)t_n + S_0^{-1}m_0) + \square$$

$$\therefore S_N^{-1} = \beta \Phi^T \Phi + S_0^{-1}$$

$$m_N S_N^{-1} = S_0^{-1} m_0 + \beta \Phi^T \mathbf{t}$$

3.11

$\sigma_N^2(x) = \frac{1}{\beta} + \phi(x)^T S_N \phi(x)$: predictive dist 분산.

(S_N : posterior의 분산). (N : 데이터 개수)

Showing that $\phi(x)^T S_{N+1} \phi(x) \leq \phi(x)^T S_N \phi(x)$

3.7 문제에서 $S_N^{-1} = \beta \Phi^T \Phi + S_0^{-1}$ 임을 보였다.

$$(S_{N+1}^{-1})^{-1} = (\beta \Phi^T \Phi + S_0^{-1})^{-1} = (\beta \phi(x)\phi(x)^T + S_N^{-1})^{-1}$$

$$= S_N - \frac{\beta S_N \phi(x)\phi(x)^T S_N}{1 + \beta \phi(x)^T S_N \phi(x)} = S_{N+1}.$$

$\Rightarrow$ $\phi(x)^T \frac{\beta \cdot S_N \phi(x)\phi(x)^T S_N}{1 + \beta \phi(x)^T S_N \phi(x)} \phi(x) \geq 0$ 임을 보이면 된다.

$\Rightarrow$ AA^T 형태이므로 n.n.d.

3.16.

Marginal likelihood function $p(\underset{\sim}{t} \mid \alpha, \beta) = \int p(t \mid w, \beta)\, p(w \mid \alpha, \beta)\, dw$

From (3.11) $p(\underset{\sim}{t} \mid X, w, \beta) = \prod_{n=1}^{N} \mathcal{N}(t_n \mid w^T \phi(x_n), \beta^{-1}) \Rightarrow \log p(\underset{\sim}{t} \mid X, w, \beta) = \frac{N}{2}\log\beta - \frac{N}{2}\log(2\pi) - \beta E_D(w)$

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

(3.52) Gaussian prior ; $p(w \mid \alpha) = \mathcal{N}(w \mid 0, \alpha^{-1} I)$

evidence function (= marginal likelihood function) is

$$p(\underset{\sim}{t} \mid \alpha, \beta) = \int \prod_{n=1}^{N} \mathcal{N}(t_n \mid w^T\phi(x_n), \beta^{-1}) \cdot \mathcal{N}(w \mid 0, \alpha^{-1} I),$$

$$= \left(\frac{\beta}{2\pi}\right)^{\frac{N}{2}} \cdot \left(\frac{\alpha}{2\pi}\right)^{\frac{M}{2}} \cdot \int \exp\{-E(w)\}\, dw.$$

(M 은 w의 dim)

$$E(w) = \beta E_D(w) + \alpha E_W(w)$$

$$= \frac{\beta}{2}\|t - \Phi w\|^2 + \frac{\alpha}{2} w^T w.$$

$$= \underbrace{\frac{\beta}{2}\|t - \Phi m_N\|^2 + \frac{\alpha}{2} m_N^T m_N}_{= E(m_N)} + \frac{1}{2}(w - m_N)^T A (w - m_N)$$

where $A = \alpha I + \beta \Phi^T \Phi$

$m_N = \beta A^{-1} \Phi^T t$

$$\therefore \log p(\underset{\sim}{t} \mid \alpha, \beta) = \underbrace{\frac{N}{2}\log\beta + \frac{M}{2}\log\alpha - \left(\frac{N}{2} + \frac{M}{2}\right)\log 2\pi}_{} + \log\int \exp(-E(w))\, dw$$

$$= \qquad \qquad + \log[\exp(-E(m_N))]$$

$$+ \log \underbrace{\int \exp\left(-\frac{1}{2}(w - m_N)^T A (w - m_N)\right)}_{= (2\pi)^{M/2} \cdot |A|^{-\frac{1}{2}}}$$

3.20.

Maximize (3.86) with respect to α.

$$\frac{\partial \log p(\mathbf{t} \mid \alpha, \beta)}{\partial \alpha} = \frac{M}{2} \cdot \frac{1}{\alpha} - \frac{\partial E(m_N)}{\partial \alpha} - \frac{1}{2} \frac{\partial \log |A|}{\partial \alpha}$$

using eigenvector equation,

$A = \alpha I + \beta \Phi^T \Phi$, $(\beta \Phi^T \Phi) u_i = \lambda_i u_i \Rightarrow$ A의 고유값은 $\lambda_i + \alpha$

$$\therefore \frac{1}{2} \frac{\partial \log |A|}{\partial \alpha} = \frac{1}{2} \cdot \frac{\partial}{\partial \alpha} \log \prod_i^M (\lambda_i + \alpha) = \frac{1}{2} \sum \frac{1}{\lambda_i + \alpha}$$

$$\frac{\partial E(m_N)}{\partial \alpha} = \frac{\partial}{\partial \alpha} \frac{\beta}{2} \| t - \Phi m_N \|^2 + \frac{\alpha}{2} m_N^T m_N = \frac{1}{2} m_N^T m_N$$

$$\Rightarrow \frac{\partial \log p(\mathbf{t} \mid \alpha, \beta)}{\partial \alpha} = \frac{M}{2\alpha} - \frac{1}{2} m_N^T m_N - \frac{1}{2} \sum \frac{1}{\lambda_i + \alpha} = 0$$ 을

만족하는 α를 찾으면 (3.86)을 극대화

즉 $$\alpha m_N^T m_N = M - \sum_{i=1}^{M} \frac{\alpha}{\lambda_i + \alpha} = \sum \frac{\lambda_i}{\lambda_i + \alpha}$$