Part 1 :

1. we got training data set $\underline{X} = [\underline{x}_1, \dots, \underline{x}_N]$

target data set $T = [\underline{t}_1, \dots, \underline{t}_N]$,

from 課本 P.309

we can get the conditional distribution of t_{N+1} for a test input vector x_{N+1}

$$p(t_{N+1} | T) = \mathcal{N}(t_{N+1} | m(x_{N+1}), \sigma^2(x_{N+1}))$$

where $m(x_{N+1}) = \underline{k}^T C_N^{-1} T$ from (6.66)

$\sigma^2(x_{N+1}) = c - \underline{k}^T C_N^{-1} \underline{k}$ from (6.67) ,

and C_N is defined from (6.62)

$$C(x_n, x_m) = k(x_n, x_m) + \beta^{-1} \delta_{nm} ,$$

so when σ^2 is univariate ,

$$C_N = k(x) + \beta^{-1} \delta ,$$

when σ^2 is multivariate ,

$$C_N = k(x_n, x_m) + \beta^{-1} \delta_{nm}$$

2. $p(\underline{t}|\underline{X},\underline{\alpha},\beta) = \int \underline{p(\underline{t}|\underline{X},\underline{w},\beta)}\,\underline{p(\underline{w}|\underline{\alpha})}\,d\underline{w}$

from 課本(3.10), $p(\underline{t}|\underline{X},\underline{w},\beta) = \prod_{n=1}^{N} \mathcal{N}(t_n|\underline{w}^T\phi(\underline{x_n}),\beta^{-1})$

and $\because$ Gaussian distribution $\mathcal{N}(x,\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$

$\therefore \Rightarrow \prod_{n=1}^{N}\left(\frac{1}{(2\pi\beta^{-1})^{\frac{1}{2}}}\exp\left(\frac{-(t_n-w^T\phi(x_n))^2}{2\beta^{-1}}\right)\right)$

$= \left(\frac{\beta}{2\pi}\right)^{\frac{N}{2}}\int \exp(-E_D(\underline{w}))dw$

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

from (7.80), $p(\underline{w}|\underline{\alpha}) = \prod_{i=1}^{M}\mathcal{N}(\omega_i|0,\alpha_i^{-1})$

$= \frac{1}{(2\pi)^{\frac{M}{2}}}\prod_{i=1}^{M}\alpha_i^{\frac{1}{2}}\exp\left(-\frac{1}{2}\sum_{i=1}^{M}\frac{(w_i)^2}{\alpha_i^{-1}}\right)$

$\therefore p(\underline{t}|\underline{X},\underline{\alpha},\beta) = \int p(\underline{t}|\underline{X},\underline{w},\beta)\,p(\underline{w}|\underline{\alpha})\,d\underline{w}$

$= \left(\frac{\beta}{2\pi}\right)^{\frac{N}{2}}\left(\frac{1}{2\pi}\right)^{\frac{M}{2}}\prod_{i=1}^{M}\alpha_i^{\frac{1}{2}}\int \underline{\exp\left(-E_D(\underline{w})-\frac{1}{2}\sum_{i=1}^{M}\alpha_i w_i^2\right)}\,d\underline{w}$

where $\left(-E_D(\underline{w})-\frac{1}{2}\sum_{i=1}^{M}\alpha_i w_i^2\right) = -\left(E_D(\underline{w})+\frac{1}{2}\sum_{i=1}^{M}\alpha_i w_i^2\right)$

$= -\left(\frac{\beta}{2}\sum_{n=1}^{N}\{t_n-\underline{w}^T\phi(x_n)\}^2+\frac{1}{2}\sum_{i=1}^{M}\alpha_i w_i^2\right)$

$= -\left(\frac{\beta}{2}\|\underline{t}-\Phi\underline{w}\|^2+\frac{1}{2}\underline{w}^TA\underline{w}\right)$,

let $\Phi = \phi_i(\underline{x_n})$, $A = diag(\alpha)$

from (7.82)、(7.83), $\underline{m} = \beta\Sigma\Phi^T\underline{t}$, $\Sigma = (A+\beta\Phi^T\Phi)^{-1}$,

$-\left(\frac{\beta}{2}\|t-\Phi\underline{w}\|^2+\frac{1}{2}\underline{w}^TA\underline{w}\right) = -\left(\frac{1}{2}(\underline{w}-\underline{m})^T\Sigma^{-1}(\underline{w}-\underline{m})+\frac{1}{2}(\beta\underline{t}^Tt-\underline{m}^T\Sigma^{-1}\underline{m})\right)$

$$\therefore \int \exp\left[-\left(\tfrac{1}{2}(\underline{w}-\underline{m})^T\underline{\Sigma}^{-1}(\underline{w}-\underline{m}) + \tfrac{1}{2}(\beta\underline{t}^T t - \underline{m}\underline{\Sigma}^{-1}\underline{m})\right]\right] dw$$

$$= \exp\left[-\left(\tfrac{1}{2}(\beta\underline{t}^T\underline{t} - \underline{m}^T\underline{\Sigma}^{-1}\underline{m})\right)\right](2\pi)^{\frac{M}{2}}|\Sigma|^{\frac{1}{2}}$$，代入 $\underline{m}$ 和 $\underline{\Sigma}$

$$= \exp\left[-\left(\tfrac{1}{2}(\beta\underline{t}^T\underline{t} - \beta\underline{t}^T\Phi\Sigma\Phi^T\underline{t}\beta)\right)\right](2\pi)^{\frac{M}{2}}|\Sigma|^{\frac{1}{2}}$$

$$= \exp\left[-\left(\tfrac{1}{2}\underline{t}^T(\beta I - \beta\Phi\Sigma\Phi^T\beta)\underline{t}\right)\right](2\pi)^{\frac{M}{2}}|\Sigma|^{\frac{1}{2}}$$

$$= \exp\left[-\left(\tfrac{1}{2}\underline{t}^T C^{-1}\underline{t}\right)\right](2\pi)^{\frac{M}{2}}|\Sigma|^{\frac{1}{2}}$$, where $C = \beta^{-1}I + \Phi A^{-1}\Phi^T$

from (7.86)

$$\therefore p(\underline{t}|\underline{X},\underline{\alpha},\beta) = \left(\frac{\beta}{2\pi}\right)^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}\prod_{m=1}^{M}\alpha_i^{\frac{1}{2}}\exp\left[-\left(\tfrac{1}{2}\underline{t}^T C^{-1}\underline{t}\right)\right]$$

$$\therefore \ln p(\underline{t}|\underline{X},\underline{\alpha},\beta) = -\tfrac{1}{2}\{N\ln(2\pi) + \ln|C| + \underline{t}^T C^{-1}\underline{t}\}$$

#

3. Show that M distinct random variables have $2^{M(M-1)/2}$ distinct indirected graphs.

if 0 edges, number of graphs : $C^{\frac{M(M-1)}{2}}_{0} = 1$

elif 1 edges, number of graphs : $C^{\frac{M(M-1)}{2}}_{1}$

elif 2 edges, number of graphs : $C^{\frac{M(M-1)}{2}}_{2}$

$\vdots$

elif $\frac{M(M-1)}{2}$ edges, number of graphs : $C^{\frac{M(M-1)}{2}}_{\frac{M(M-1)}{2}}$

so total number of graphs are :

$$C^{\frac{M(M-1)}{2}}_{0} + C^{\frac{M(M-1)}{2}}_{1} + C^{\frac{M(M-1)}{2}}_{2} + \dots + C^{\frac{M(M-1)}{2}}_{\frac{M(M-1)}{2}},$$

and using Binomial expansion theorem,

$$C^{\frac{M(M-1)}{2}}_{0} 1^{\left(\frac{M(M-1)}{2}-0\right)} 1^{0} + C^{\frac{M(M-1)}{2}}_{1} 1^{\left(\frac{M(M-1)}{2}-1\right)} 1^{1} + \dots + C^{\frac{M(M-1)}{2}}_{\frac{M(M-1)}{2}} 1^{0} 1^{\left(\frac{M(M-1)}{2}-0\right)}$$

$$= (1+1)^{\frac{M(M-1)}{2}}$$

$$= 2^{\frac{M(M-1)}{2}}$$

#

4. Show α_i^{new} and $(\beta^{new})^{-1}$ in Section 7 is formally equivalent in section 9.

a_i^{new} :

$$\alpha_i^{new} = \frac{\gamma_i}{m_i^2} \quad , \quad \gamma_i = 1-\alpha_i \Sigma_{ii} \quad \text{from } (7.87),(7.89)$$

$$\Rightarrow \alpha_i^{new} = \frac{1-\alpha_i^{new}\Sigma_{ii}}{m_i^2} \Rightarrow \alpha_i^{new} m_i^2 = 1-\alpha_i^{new}\Sigma_{ii}$$

$$\Rightarrow \alpha_i^{new}(m_i^2+\Sigma_{ii}) = 1$$

$$\Rightarrow \alpha_i^{new} = \frac{1}{m_i^2+\Sigma_{ii}}$$

#

$(\beta^{new})^{-1}$:

$$(\beta^{new})^{-1} = \frac{\|\underline{t}-\Phi \underline{m}_N\|^2+\beta^{-1}\Sigma_i\gamma_i}{N}$$

$$= \frac{\|\underline{t}-\Phi \underline{m}_N\|^2}{N} + \frac{\Sigma_i\gamma_i}{\beta^{new}N}$$

$$\Rightarrow (\beta^{new})^{-1} - \frac{\Sigma_i\gamma_i}{\beta^{new}N} = \frac{\|\underline{t}-\Phi \underline{m}_N\|^2}{N}$$

$$\Rightarrow \frac{N-\Sigma_i\gamma_i}{\beta^{new}N} = \frac{\|\underline{t}-\Phi \underline{m}_N\|^2}{N}$$

$$\Rightarrow (\beta^{new})^{-1} = \frac{\cancel{N}}{N-\Sigma_i\gamma_i} \cdot \frac{\|\underline{t}-\Phi \underline{m}_N\|^2}{\cancel{N}}$$

$$= \frac{\|\underline{t}-\Phi \underline{m}_N\|^2}{N-\Sigma_i\gamma_i}$$

#