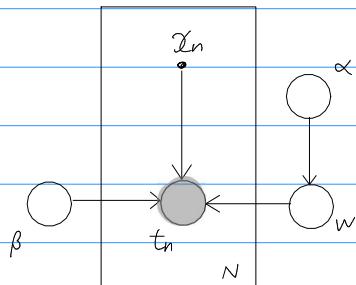


10.26



二元モデルに対する同時分布の分解は

$$p(\pi, w, \alpha, \beta | X) = p(\pi | w, \beta, X) p(w | \alpha) p(\alpha) p(\beta) \dots \quad (1)$$

とします。

w, α, β の尤度関数と事前分布を

$$p(\pi | w, \beta, X) = \prod_{n=1}^N N(t_n | w^\top \phi_n, \beta^{-1})$$

$$p(w | \alpha) = N(w | 0, \alpha^{-1} I)$$

$$p(\alpha) = \text{Gam}(\alpha | a_0, b_0)$$

$$p(\beta) = \text{Gam}(\beta | c_0, d_0)$$

とします。

w, α, β の事後分布 $p(w, \alpha, \beta | \pi, X)$ の近似を $g(w, \alpha, \beta)$ とし

$$g(w, \alpha, \beta) = g(w) g(\alpha, \beta)$$

と分解できることと仮定する。

ここで“モデル” $\alpha \perp\!\!\!\perp \beta | 0$ だから、さらに

$$g(w, \alpha, \beta) = g(w) g(\alpha, \beta) = g(w) g(\alpha) g(\beta)$$

と分解できます。

$(g(\alpha), g(\beta), g(w))$ の更新式を求める

(10.9) 式)

$$\begin{aligned}
\ln g(\alpha) &= E_{w,\beta} [\ln p(t, w, \alpha, \beta | X)] \\
&= E_{w,\beta} [\ln p(t | w, \beta, X) p(w | \alpha) p(\alpha) p(\beta)] \\
&= E_{w,\beta} [\ln p(t | w, \beta, X)] + E_w [\ln p(w | \alpha)] + \ln p(\alpha) + E_\beta [\ln p(\beta)] \\
&= E_w [\ln p(w | \alpha)] + \ln p(\alpha) + C \quad \leftarrow \alpha \text{ に依存する項} \\
&= E_w [\ln N(w | \alpha_0, \alpha^{-1} I)] + \ln \text{Gam}(\alpha | a_0, b_0) + C \\
&= E_w [\ln \frac{1}{(2\pi)^M} \frac{1}{(\sqrt{w})^M} \exp(-\frac{\alpha}{2} w^T w)] + \ln \frac{1}{\Gamma(a_0)} b_0^{a_0} \alpha^{a_0-1} e^{-b_0 \alpha} + C \\
&= \frac{M}{2} \ln \alpha - \frac{\alpha}{2} E[w^T w] + (a_0-1) \ln \alpha - b_0 \alpha + C \\
&= (\frac{M}{2} + a_0 - 1) \ln \alpha - (\frac{1}{2} E[w^T w] + b_0) \alpha + C
\end{aligned}$$

二の式)

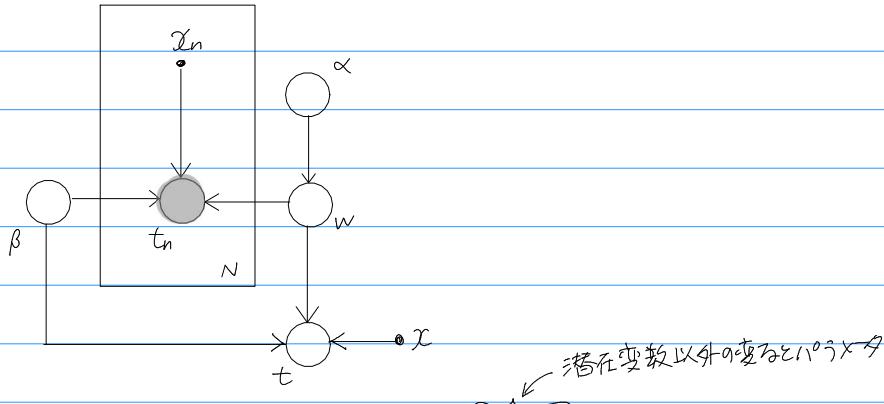
$$g(\alpha) = \text{Gam}(\alpha | a_N, b_N), a_N = \frac{M}{2} + a_0, b_N = \frac{1}{2} E[w^T w] + b_0$$

を得る

同様 (10.9) 式)

$$\begin{aligned}
\ln g(\beta) &= E_{w,\alpha} [\ln p(t, w, \alpha, \beta | X)] \\
&= E_{w,\alpha} [\ln p(t | w, \beta, X) p(w | \alpha) p(\alpha) p(\beta)] \\
&= E_{w,\alpha} [\ln p(t | w, \beta, X)] + \ln p(\beta) + C \quad \leftarrow \beta \text{ に依存する項} \\
&= E_w [\ln \prod_{n=1}^N N(t_n | w^T \phi_n, \beta^{-1})] + \ln \text{Gam}(\beta | c_0, d_0) + C \\
&= \sum_{n=1}^N E_w [\ln \frac{1}{(2\pi)^{\frac{1}{2}}} \frac{1}{(\beta)^{\frac{1}{2}}} \exp\{-\frac{\beta}{2} (t_n - w^T \phi_n)^2\}] + \ln \frac{1}{\Gamma(c_0)} d_0^{c_0-1} \beta^{c_0-1} e^{-d_0 \beta} + C \\
&= \frac{N}{2} \ln \frac{\beta}{2\pi} + \sum_{n=1}^N (-\frac{\beta}{2}) E_w [(t_n - w^T \phi_n)^2] + (c_0-1) \ln \beta - d_0 \beta + C
\end{aligned}$$

(予測分布を求める)



ここで求めた予測分布を $p(t|x, \beta, X)$ とすると

$$\begin{aligned}
 p(t|x, \beta, X) &= \int p(t, w, \beta, \alpha | x, t, X) dw d\alpha d\beta \\
 &= \int p(t|w, \beta, \alpha, x, t, X) p(w, \beta, \alpha | x, t, X) dw d\alpha d\beta \\
 &= \int p(t|w, \beta, x) p(w, \beta, \alpha | t, X) dw d\alpha d\beta \quad \leftarrow t \perp\!\!\!\perp (t, \alpha | x) \mid (w, \beta, x) \text{ 事後} \\
 &\quad \leftarrow w, \beta, \alpha \text{ の事後分布の並び} \quad p(t|w, \beta, \alpha, x, t, X) = p(t|w, \beta, x) \\
 &\simeq \int p(t|w, \beta, x) g(w, \beta, \alpha) dw d\alpha d\beta \quad (w, \beta, \alpha) \perp\!\!\!\perp x \mid (t, X) \text{ 事後} \\
 &\quad \leftarrow p(w, \beta, \alpha | x, t, X) = p(w, \beta, \alpha | t, X) \\
 &= \int p(t|w, \beta, x) g(w) g(\beta) g(\alpha) dw d\alpha d\beta \\
 &= \int p(t|w, \beta, x) g(w) g(\beta) dw d\beta \\
 &= \int N(t | w^T \phi(x), \beta^{-1}) N(w | m_N, S_N) \text{Gam}(\beta | c_N, d_N) dw d\beta \\
 &\quad \leftarrow (3.8)(3.3) \\
 &= \int N(t | w^T \phi(x), \beta^{-1}) \text{Gam}(\beta | c_N, d_N) d\beta N(w | m_N, S_N) dw \\
 &\rightarrow \int N(t | w^T \phi(x), E[\beta]^{-1}) N(w | m_N, S_N) dw \quad (N \rightarrow \infty) \quad \leftarrow \int_0^\infty N(t | w^T \phi(x), \beta^{-1}) \text{Gam}(\beta | c_N, d_N) d\beta \quad \leftarrow \beta = \gamma^{-1} \text{ 事後} \\
 &\quad \leftarrow \frac{d\beta}{d\gamma} = \frac{-1}{\gamma^2} d\gamma \quad \leftarrow \beta \Big|_{0}^{\infty} \rightarrow \infty \\
 &= N(t | m_N^T \phi(x), E[\beta]^{-1} + \phi(x)^T S_N \phi(x)) \quad \leftarrow \int_0^\infty N(t | w^T \phi(x), \gamma) \frac{1}{\Gamma(c_N)} \frac{c_N^{c_N}}{d_N^{d_N}} \gamma^{-(c_N+1)} \exp(-\frac{d_N}{\gamma}) d\gamma \\
 &\quad \leftarrow N(t | w^T \phi(x), \gamma) \text{InverseGamma}(\gamma | c_N, d_N) d\gamma \\
 &\rightarrow \int_0^\infty N(t | w^T \phi(x), \gamma) \delta(\frac{dw}{c_N}) d\gamma \quad (N \rightarrow \infty) \quad \leftarrow \int_0^\infty N(t | w^T \phi(x), \gamma) \delta(\frac{dw}{c_N}) d\gamma \quad (N \rightarrow \infty) \\
 &= N(t | w^T \phi(x), \frac{dw}{c_N}) = N(t | w^T \phi(x), E[\beta]^{-1})
 \end{aligned}$$

を得る。(ただし \$N\$ が大きいたとき)

(\$N\$ が小さいときは?)

$\text{Gam}(y | \gamma, \theta)$

$$\begin{aligned}
 \text{mode} &= \frac{dy}{d\gamma} = \frac{d\gamma}{d\theta} \rightarrow \frac{d\gamma}{d\theta} (N \rightarrow \infty) \quad \leftarrow \left\{ \begin{array}{l} c_N = \frac{N}{2} + c_0 \\ d_N = N - c_N \end{array} \right. \\
 &\quad \leftarrow d_N \neq 0 \Rightarrow \text{mode} \neq 0 \quad \leftarrow \text{mode } 0, \text{var } 0 \text{ に近づく} \\
 \text{var} &= \frac{d\gamma^2}{d\theta^2} = \frac{d\theta^2}{d\gamma^2} \rightarrow \frac{d\theta^2}{d\gamma^2} (N \rightarrow \infty) \quad \leftarrow \text{mode } 0, \text{var } 0 \text{ に近づく} \\
 &\quad \leftarrow \text{mode } 0, \text{var } 0 \text{ に近づく}
 \end{aligned}$$

$$\text{InverseGamma}(\gamma | c_N, d_N) \rightarrow \delta(\frac{dw}{c_N}) (N \rightarrow \infty)$$

(変分下界を求める)

← 10.3×→ Xには省略

$$\mathcal{L}(g) = \int g(w, \alpha, \beta) \ln \frac{p(t|w, \alpha, \beta)}{q(w, \alpha, \beta)} dw d\alpha d\beta$$

$$= E_{w, \alpha, \beta} [\ln p(t|w, \alpha, \beta)] - E_{w, \alpha, \beta} [\ln q(w, \alpha, \beta)]$$

$$= E_{w, \alpha, \beta} [\ln p(t|w, \beta) p(w|\alpha) p(\alpha) p(\beta)] - E_{w, \alpha, \beta} [\ln q(w) q(\alpha) q(\beta)]$$

$$= E_{w, \beta} [\ln p(t|w, \beta)] + E_{w, \alpha} [\ln p(w|\alpha)] + E_{\alpha} [\ln p(\alpha)] + E_{\beta} [\ln p(\beta)]$$

$$- E_w [\ln q(w)] - E_{\alpha} [\ln q(\alpha)] - E_{\beta} [\ln q(\beta)]$$

各期待値を求める

$$E_{w, \beta} [\ln p(t|w, \beta)] = E_{w, \beta} \left[\ln \frac{N}{N} N(t_n | w^T \phi_n, \beta^{-1}) \right]$$

$$= \sum_{n=1}^N E_{w, \beta} [\ln N(t_n | w^T \phi_n, \beta^{-1})]$$

$$= \sum_{n=1}^N E_{w, \beta} \left[\ln \frac{1}{(2\pi)^{\frac{1}{2}}} \frac{1}{(\beta^{-1})^{\frac{1}{2}}} \exp \left\{ -\frac{\beta}{2} (t_n - w^T \phi_n)^2 \right\} \right]$$

$$= -\frac{N}{2} \ln(2\pi) + \frac{N}{2} E_{\beta} [\ln \beta] - \frac{1}{2} \sum_{n=1}^N E_{\beta} [\beta] E_w [(t_n - w^T \phi_n)^2]$$

$$= -\frac{N}{2} \ln(2\pi) + \frac{N}{2} \left\{ \text{H}(C_N) - \ln d_N \right\} - \frac{1}{2} \frac{C_N}{d_N} \sum_{n=1}^N E_w [(t_n - w^T \phi_n)^2] \quad \begin{matrix} (\text{B.30}) \\ (\text{B.27}) \end{matrix}$$

$$= -\frac{N}{2} \ln(2\pi) + \frac{N}{2} \left\{ \text{H}(C_N) - \ln d_N \right\} - \frac{1}{2} \frac{C_N}{d_N} \left\{ t^T - 2m_N^T \bar{t} + \text{Tr}(\bar{t}^T \bar{t} (m_N m_N^T + S_N)) \right\}$$

$E_{w, \alpha} [\ln p(w|\alpha)]$ は (10.109) と同じ

$\sum E_w [(t_n - w^T \phi_n)^2]$ の計算上で $g(\beta)$ を求めたときに
出でたのと同じ

$E_{\alpha} [\ln p(\alpha)]$ は (10.110) と同じ

$$E_{\beta} [\ln p(\beta)] = E_{\beta} [\ln \text{Gam}(\beta | C_0, d_0)] = E_{\beta} \left[\ln \frac{1}{\Gamma(C_0)} d_0^{C_0} \beta^{C_0-1} \exp(-\beta d_0) \right]$$

$$= -\ln P(C_0) + C_0 \ln d_0 + (C_0 - 1) E_{\beta} [\ln \beta] - d_0 E_{\beta} [\beta]$$

$$= -\ln P(C_0) + C_0 \ln d_0 + (C_0 - 1) \left(\text{H}(C_N) - \ln d_N \right) - d_0 \frac{C_N}{d_N} \quad \begin{matrix} (\text{B.30}) \\ (\text{B.27}) \end{matrix}$$

$- E_w [\ln q(w)]$ は (10.111) と同じ

$- E_{\alpha} [\ln q(\alpha)]$ は (10.112) と同じ

$$\begin{aligned}
-\mathbb{E}_\beta [\ln \vartheta(\beta)] &= -\mathbb{E}_\beta \left[\ln \frac{1}{\Gamma(c_N)} dN \beta^{c_N-1} \exp(-\beta dN) \right] \\
&= \ln \Gamma(c_N) - c_N \ln dN - (c_N-1) \mathbb{E}_\beta [\ln \beta] + dN \mathbb{E}_\beta [\beta] \\
&= \ln \Gamma(c_N) - c_N \ln dN - (c_N-1) (\psi(c_N) - \ln dN) + dN \frac{c_N}{dN} \\
&= \ln \Gamma(c_N) - (c_N-1) \psi(c_N) - \ln dN + c_N
\end{aligned}$$