14.12

t_nがベクトルの場合、t_nの条件付分布を混合ガウス分布で表わすと

$$p(t_n|\phi_n,\theta)=\sum_{k=1}^{K}\pi_k N(t_n|W_k^T\phi_n,\beta^{-1}I) \cdots (14.34)'$$

ここで、t_nがD次元ベクトル、ϕ_nがM次元ベクトルのとき W_kは $M\times D$次元の行列である。

これより尤度関数は

$$p(t_1\cdots t_N|\phi_1\cdots\phi_N,\theta)\underset{(t_n,\phi_n)はiidなので}{=}\prod_{n=1}^{N}p(t_n|\phi_n,\theta)=\prod_{n=1}^{N}\sum_{k=1}^{K}\pi_k N(t_n|W_k^T\phi_n,\beta^{-1}I)$$

となる。対数尤度関数は

$$\ln p(t_1\cdots t_N|\phi_1\cdots\phi_N,\theta)=\sum_{n=1}^{N}\ln\left(\sum_{k=1}^{K}\pi_k N(W_k^T\phi_n,\beta^{-1}I)\right) \cdots (14.35)'$$

となる。

9.2節に倣って潜在変数z_nを導入したとき、(9.10), (9.11)と同じように

$$p(z_n|\theta)=\prod_{k=1}^{K}\pi_k^{z_{nk}}$$

$$p(t_n|z_n,\phi_n,\theta)=\prod_{k=1}^{K}N(t_n|W_k^T\phi_n,\beta^{-1}I)^{z_{nk}}$$

である。これより

z_nはϕ_nに依らないので$p(z_n|\theta)=p(z_n|\phi_n,\theta)$

$$p(t_n,z_n|\phi_n,\theta)=p(z_n|\theta)p(t_n|z_n,\phi_n,\theta)=\prod_{k=1}^{K}\{\pi_k N(t_n|W_k^T\phi_n,\beta^{-1}I)\}^{z_{nk}} \cdots ①$$

①を用いて完全データ集合の尤度は

$$p(t_1\cdots t_N,z_1\cdots z_N|\phi_1\cdots\phi_N,\theta)\underset{(t_n,z_n,\phi_n)はiidなので}{=}\prod_{n=1}^{N}p(t_n,z_n|\phi_n,\theta)=\prod_{n=1}^{N}\prod_{k=1}^{K}\{\pi_k N(t_n|W_k^T\phi_n,\beta^{-1}I)\}^{z_{nk}}$$

よって完全データ集合の対数尤度は

$$\ln p(t_1\cdots t_N,z_1\cdots z_N|\phi_1\cdots\phi_N,\theta)=\sum_{n=1}^{N}\sum_{k=1}^{K}z_{nk}\ln\{\pi_k N(t_n|W_k^T\phi_n,\beta^{-1}I)\} \cdots (14.36)'$$

となる。負担率は

$$\gamma_{nk}=p(z_n=k|t_n,\phi_n,\theta^{old})=\frac{p(z_n=k,t_n|\phi_n,\theta^{old})}{p(t_n|\phi_n,\theta^{old})}=\frac{p(z_n=k,t_n|\phi_n,\theta^{old})}{\sum_{z_n}p(z_n,t_n|\phi_n,\theta^{old})}$$

$$=\frac{\left[\prod_{j=1}^{K}\{\pi_j N(t_n|W_j^T\phi_n,\beta^{-1}I)\}^{z_{nj}}\right]_{z_n=k}}{\sum_{z_n}\prod_{j=1}^{K}\{\pi_j N(t_n|W_j^T\phi_n,\beta^{-1}I)\}^{z_{nj}}}=\frac{\pi_k N(t_n|W_k^T\phi_n,\beta^{-1}I)}{\sum_{k=1}^{K}\pi_k N(t_n|W_k^T\phi_n,\beta^{-1}I)} \cdots (14.37)'$$

となる。

これらを用いてQ関数は

$$Q(\theta, \theta^{old}) = \mathop{E}_{p(z_1 \cdots z_N | t_1 \cdots t_N, \phi_1 \cdots \phi_N, \theta^{old})}[\ln p(t_1 \cdots t_N, z_1 \cdots z_N | \phi_1 \cdots \phi_N, \theta)]$$

$$= \sum_{z_1} \cdots \sum_{z_N} p(z_1 \cdots z_N | t_1 \cdots t_N, \phi_1 \cdots \phi_N, \theta^{old}) \ln p(t_1 \cdots t_N, z_1 \cdots z_N | \phi_1 \cdots \phi_N, \theta)$$

$$= \sum_{z_1} \cdots \sum_{z_N} p(z_1 \cdots z_N | t_1 \cdots t_N, \phi_1 \cdots \phi_N, \theta^{old}) \sum_{n=1}^{N} \sum_{k=1}^{K} z_{nk} \ln \{ \pi_k N(t_n | W_k^T \phi_n, \beta^{-1} I) \} \quad (14.36)'$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} \sum_{z_1} \cdots \sum_{z_N} p(z_1 \cdots z_N | t_1 \cdots t_N, \phi_1 \cdots \phi_N, \theta^{old}) z_{nk} \ln \{ \pi_k N(t_n | W_k^T \phi_n, \beta^{-1} I) \}$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma_{nk} \ln \{ \pi_k N(t_n | W_k^T \phi_n, \beta^{-1} I) \} \longleftarrow \gamma_{nk} = \mathop{E}_{p(z_1 \cdots z_N | t_1 \cdots t_N, \phi_1 \cdots \phi_N, \theta^{old})}[z_{nk}] = \sum_{z_1} \cdots \sum_{z_N} p(z_1 \cdots z_N | t_1 \cdots t_N, \phi_1 \cdots \phi_N, \theta^{old}) z_{nk}$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma_{nk} \{ \ln \pi_k + \ln N(t_n | W_k^T \phi_n, \beta^{-1} I) \}$$

となる。

(π_kについての最適化)

π_kについての項は、本文と同じなので、π_kの再推定式は(14.38)と同じ

(W_kについての最適化)

Q関数でW_kを含む項を分りすると

$$Q(\theta, \theta^{old}) = \sum_{n=1}^{N} \gamma_{nk} \ln N(t_n | W_k^T \phi_n, \beta^{-1} I) + C, \quad C: W_k\text{を含まない項}$$

$$= \sum_{n=1}^{N} \gamma_{nk} \ln \frac{1}{|2\pi|^{\frac{D}{2}}} \frac{1}{|\beta^{-1}|^{\frac{D}{2}}} \exp \{ -\frac{\beta}{2} (t_n - W_k^T \phi_n)^T (t_n - W_k^T \phi_n) \} + C$$

$$= \sum_{n=1}^{N} \gamma_{nk} \{ -\frac{\beta}{2} (t_n - W_k^T \phi_n)^T (t_n - W_k^T \phi_n) \} + C \quad \cdots (14.39)'$$

ここで

$$\frac{\partial \mathbf{u}^T\mathbf{u}}{\partial x} = \frac{\partial (u_1^2+u_2^2)}{\partial x} = 2u_1\frac{\partial u_1}{\partial x} + 2u_2\frac{\partial u_2}{\partial x} = 2\mathbf{u}^T\frac{\partial \mathbf{u}}{\partial x}$$

$$\frac{\partial (\mathbf{t}_n - W_k^T\phi_n)^T(\mathbf{t}_n - W_k^T\phi_n)}{\partial W_{kij}} = 2(\mathbf{t}_n - W_k^T\phi_n)^T\frac{\partial(\mathbf{t}_n - W_k^T\phi_n)}{\partial W_{kij}}$$

$$= 2(\mathbf{t}_n - W_k^T\phi_n)^T(-J^{ji}\phi_n)$$

$$\frac{\partial W^T\phi}{\partial W_{12}} = \frac{\partial}{\partial W_{12}}\begin{pmatrix} w_{11} & w_{21} \\ w_{12} & w_{22}\end{pmatrix}\begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix} = J^{21}\begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix}$$

よって

$$\frac{\partial W_k^T\phi_n}{\partial W_{kij}} = J^{ji}\phi_n$$

$$= -2\{(\mathbf{t}_n - W_k^T\phi_n)\phi_n^T\}_{ji}$$

$$\mathbf{a}^T J^{21}\mathbf{b} = (a_1\ a_2)\begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix}\begin{pmatrix}b_1\\ b_2\end{pmatrix} = (a_1\ a_2)\begin{pmatrix}0\\ b_1\end{pmatrix} = a_2b_1$$

また

$$\mathbf{a}\mathbf{b}^T = \begin{pmatrix}a_1\\ a_2\end{pmatrix}(b_1\ b_2) = \begin{pmatrix}a_1b_1 & a_1b_2\\ a_2b_1 & a_2b_2\end{pmatrix}$$

よって

$$\mathbf{a}^T J^{21}\mathbf{b} = (\mathbf{a}\mathbf{b}^T)_{21}$$

なので

$$\frac{\partial Q}{\partial W_{kij}} = \sum_{n=1}^{N}\gamma_{nk}\left[-\frac{\beta}{2}(-2)\{(\mathbf{t}_n - W_k^T\phi_n)\phi_n^T\}_{ji}\right]$$

$$= \sum_{n=1}^{N}\gamma_{nk}\,\beta\{(\mathbf{t}_n - W_k^T\phi_n)\phi_n^T\}_{ji} = \sum_{n=1}^{N}\gamma_{nk}\,\beta\{\phi_n(\mathbf{t}_n - W_k^T\phi_n)^T\}_{ij}$$

よって

$$\frac{\partial Q}{\partial W_k} = \sum_{n=1}^{N}\gamma_{nk}\,\beta\,\phi_n(\mathbf{t}_n - W_k^T\phi_n)^T$$

となる。これより Q を最大化する W_k は

$$O = \frac{\partial Q}{\partial W_k} = \sum_{n=1}^{N}\gamma_{nk}\,\beta\,\phi_n(\mathbf{t}_n - W_k^T\phi_n)^T$$

$$\therefore\ O = \sum_{n=1}^{N}\gamma_{nk}\,\phi_n(\mathbf{t}_n - W_k^T\phi_n)^T \quad \cdots (14.40)'$$

で与えられる。

ここで $\Phi = \begin{pmatrix}\phi_1^T\\ \vdots\\ \phi_N^T\end{pmatrix}$, $T = \begin{pmatrix}\mathbf{t}_1^T\\ \vdots\\ \mathbf{t}_N^T\end{pmatrix}$, $R_k = \begin{pmatrix}\gamma_{1k} & & 0\\ & \ddots & \\ 0 & & \gamma_{Nk}\end{pmatrix}$ とすると

$$\Phi^T R_k(T - \Phi W_k) = (\phi_1 \cdots \phi_N)\begin{pmatrix}\gamma_{1k} & & 0\\ & \ddots & \\ 0 & & \gamma_{Nk}\end{pmatrix}\left\{\begin{pmatrix}\mathbf{t}_1^T\\ \vdots\\ \mathbf{t}_N^T\end{pmatrix} - \begin{pmatrix}\phi_1^T\\ \vdots\\ \phi_N^T\end{pmatrix}W_k\right\}$$

$$= \sum_{n=1}^{N}\phi_n\gamma_{nk}(\mathbf{t}_n^T - \phi_n^T W_k)$$

$$= \sum_{n=1}^{N}\phi_n\gamma_{nk}(\mathbf{t}_n - W_k\phi_n)^T$$

なので、(14.40)′ は

$$O = \Phi^T R_k(T - \Phi W_k) \quad \cdots (14.41)'$$

となる。

これより

$$0 = \Phi^T R_k \mathrm{T} - \Phi^T R_k \Phi W_k$$

$$\therefore\ W_k = (\Phi^T R_k \Phi)^{-1} \Phi^T R_k \mathrm{T} \quad \cdots (14.42)'$$

を得る。

(βについて最適化)

Q関数でβを含む項を分りすると

$$Q(\theta, \theta^{old}) = \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma_{nk} \left[\ln \frac{1}{|2\pi|^{\frac{D}{2}}} \frac{1}{|\beta^{-1}|^{\frac{D}{2}}} \exp\left\{ -\frac{\beta}{2} (t_n - W_k^T \phi_n)^T (t_n - W_k^T \phi_n) \right\} \right] + C$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma_{nk} \left\{ \frac{D}{2} \ln \beta - \frac{\beta}{2} (t_n - W_k^T \phi_n)^T (t_n - W_k^T \phi_n) \right\} + C,$$ Cはβを含まない項 $\cdots (14.43)'$

Qを最大化するβは

$$0 = \frac{\partial Q}{\partial \beta} = \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma_{nk} \left\{ \frac{D}{2} \frac{1}{\beta} - \frac{1}{2} (t_n - W_k^T \phi_n)^T (t_n - W_k^T \phi_n) \right\}$$

これより $\sum_{k=1}^{K} \gamma_{nk} = \sum_{k=1}^{K} p(z_n = k \mid t_n, \phi_n, \theta) = 1$ より

$$0 = \frac{D}{\beta} N - \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma_{nk} (t_n - W_k^T \phi_n)^T (t_n - W_k^T \phi_n)$$

$$\therefore\ \frac{1}{\beta} = \frac{1}{DN} \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma_{nk} (t_n - W_k^T \phi_n)^2 \quad \cdots (14.44)'$$

を得る。