6.27

ノートのタイトル 2016/02/13

問題文の訳がおかしい。

(誤) (6.91), (6.92), (6.94) と対数尤度の勾配を用いて表わせ

(正) 対数尤度の勾配の項に使う (6.91), (6.92), (6.94) を導け

(6.90) の導出

$$p(\mathbf{t}_N|\theta) = \int p(\mathbf{t}_N|\mathbf{a}_N)p(\mathbf{a}_N|\theta)\,d\mathbf{a}_N \quad \cdots (6.89)$$

において、積分の中を $f(\mathbf{a}_N)$ とおく

$$f(\mathbf{a}_N) = p(\mathbf{t}_N|\mathbf{a}_N)p(\mathbf{a}_N|\theta)$$

ここで

$$p(\mathbf{a}_N) = \frac{1}{Z}f(\mathbf{a}_N)$$

のラプラス近似を考えると (4.135) より

$$Z = \int f(\mathbf{a}_N)\,d\mathbf{a}_N \simeq f(\mathbf{a}_N^*)\frac{(2\pi)^{N/2}}{|A|^{1/2}}$$

ここで (6.80) より

$$\Psi(\mathbf{a}_N) = \ln f(\mathbf{a}_N)$$

なので

$$f(\mathbf{a}_N^*) = \exp(\Psi(\mathbf{a}_N^*))$$

また

$$A = -\nabla\nabla \ln f(\mathbf{a}_N) = -\nabla\nabla\Psi(\mathbf{a}_N) \overset{(6.85)}{=} W_N + C_N^{-1}$$

これらより

$$Z \simeq \exp(\Psi(\mathbf{a}_N^*))\frac{(2\pi)^{N/2}}{|W_N + C_N^{-1}|}$$

を得る

$$Z = \int f(\mathbf{a}_N)\,d\mathbf{a}_N = \int p(\mathbf{t}_N|\mathbf{a}_N)p(\mathbf{a}_N|\theta)\,d\mathbf{a}_N = p(\mathbf{t}_N|\theta)$$

なので

$$\ln p(\mathbf{t}_N|\theta) = \ln Z = \Psi(\mathbf{a}_N^*) - \frac{1}{2}\ln|W_N + C_N^{-1}| + \frac{N}{2}\ln(2\pi) \quad \cdots (6.90)$$

を得る。

(6.91) の導出

(6.90) に (6.80) を入れて

$$\ln p(\mathbf{t}_N|\theta) = -\frac{1}{2}\mathbf{a}_N^{*} C_N^{-1}\mathbf{a}_N^{*} - \frac{1}{2}\ln|C_N| + \mathbf{t}_N^{T}\mathbf{a}_N^{*} - \sum \ln(1+e^{a_n^{*}}) - \frac{1}{2}\ln|W_N + C_N^{-1}|$$

を得る。 $\mathbf{a}_N^{*}$ を固定して θ_j で偏微分すると

$\frac{\partial A^{-1}}{\partial x} = -A^{-1}\frac{\partial A}{\partial x}A^{-1}$

a_N^{*} 固定した 偏微分

$$\frac{\partial}{\partial\theta_j}\ln p(\mathbf{t}_N|\theta) = -\frac{1}{2}\mathbf{a}_N^{*} C_N^{-1}\frac{\partial C_N}{\partial\theta_j}C_N^{-1}\mathbf{a}_N^{*} - \frac{1}{2}\mathrm{Tr}\left(C_N^{-1}\frac{\partial C_N}{\partial\theta_j}\right) - \frac{1}{2}\mathrm{Tr}\left((W_N + C_N^{-1})^{-1}(-1)C_N^{-1}\frac{\partial C_N}{\partial\theta_j}C_N^{-1}\right)$$

$\mathrm{Tr}(AB)=\mathrm{Tr}(BA)$

$$= -\frac{1}{2}\mathbf{a}_N^{*} C_N^{-1}\frac{\partial C_N}{\partial\theta_j}C_N^{-1}\mathbf{a}_N^{*} - \frac{1}{2}\mathrm{Tr}\left(C_N^{-1}\frac{\partial C_N}{\partial\theta_j}\right) + \frac{1}{2}\mathrm{Tr}\left(C_N^{-1}(W_N + C_N^{-1})^{-1}C_N^{-1}\frac{\partial C_N}{\partial\theta_j}\right)$$

$$= -\frac{1}{2}\mathbf{a}_N^{*} C_N^{-1}\frac{\partial C_N}{\partial\theta_j}C_N^{-1}\mathbf{a}_N^{*} - \frac{1}{2}\mathrm{Tr}\Big(\underbrace{C_N^{-1}\frac{\partial C_N}{\partial\theta_j} - C_N^{-1}(W_N + C_N^{-1})^{-1}C_N^{-1}\frac{\partial C_N}{\partial\theta_j}}_{①}\Big)$$

ここで

$$① = C_N^{-1}(W_N + C_N^{-1})^{-1}\left\{(W_N + C_N^{-1}) - C_N^{-1}\right\}\frac{\partial C_N}{\partial\theta_j}$$

$$= C_N^{-1}(W_N + C_N^{-1})^{-1}W_N\frac{\partial C_N}{\partial\theta_j}$$

$$= \left\{(W_N + C_N^{-1})C_N\right\}^{-1}W_N\frac{\partial C_N}{\partial\theta_j}$$

$$= (W_N C_N + I)^{-1}W_N\frac{\partial C_N}{\partial\theta_j}$$

よって

本文では $C_N W_N$ になっているがまちがっている?

$$\frac{\partial}{\partial\theta_j}\ln p(\mathbf{t}_N|\theta) = -\frac{1}{2}\mathbf{a}_N^{*} C_N^{-1}\frac{\partial C_N}{\partial\theta_j}C_N^{-1}\mathbf{a}_N^{*} - \frac{1}{2}\mathrm{Tr}\left((W_N C_N + I)^{-1}W_N\frac{\partial C_N}{\partial\theta_j}\right) \quad \cdots (6.91)$$

を得る。

(6.92) の導出

(6.90) は

$$\ln p(\mathbf{t}_N|\theta) = \Psi(\mathbf{a}_N^*) - \frac{1}{2}\ln|W_N + C_N^{-1}| + \frac{N}{2}\ln(2\pi)$$

これの $\mathbf{a}_N^*$ を通した θ_j での偏微分の項は、チェインルールを使って

$$\left\{\frac{\partial}{\partial \mathbf{a}_N^*}\ln p(\mathbf{t}_N|\theta)\right\}^T \frac{\partial \mathbf{a}_N^*}{\partial \theta_j} = \left(\frac{\partial \Psi(\mathbf{a}_N^*)}{\partial \mathbf{a}_N^*}\right)^T \frac{\partial \mathbf{a}_N^*}{\partial \theta_j} - \frac{1}{2}\left(\frac{\partial}{\partial \mathbf{a}_N^*}\ln|W_N + C_N^{-1}|\right)^T \frac{\partial \mathbf{a}_N^*}{\partial \theta_j} \quad \cdots ①$$

である。ここで

$$\frac{\partial \Psi(\mathbf{a}_N^*)}{\partial \mathbf{a}_N^*} = \left.\frac{\partial \Psi(\mathbf{a}_N)}{\partial \mathbf{a}_N}\right|_{\mathbf{a}_N=\mathbf{a}_N^*} = 0 \quad \cdots ②$$

また

$$\left(\frac{\partial}{\partial \mathbf{a}_N^*}\ln|W_N + C_N^{-1}|\right)_{n\text{成分}} = \frac{\partial}{\partial a_n^*}\ln|W_N + C_N^{-1}| \leftarrow (C.22)$$

W_N は $\mathbf{a}_N^*$ に依存する、C_N^{-1} は $\mathbf{a}_N^*$ に依存しないので $\dfrac{\partial(W_N + C_N^{-1})}{\partial a_n^*} = \dfrac{\partial W_N}{\partial a_n^*}$

$$= \mathrm{Tr}\left[(W_N + C_N^{-1})^{-1}\frac{\partial W_N}{\partial a_n^*}\right]$$

$$= \mathrm{Tr}\left[(W_N + C_N^{-1})^{-1}\frac{\partial}{\partial a_n^*}\begin{pmatrix}\sigma(a_1^*)(1-\sigma(a_1^*)) & & 0 \\ & \ddots & \\ 0 & & \sigma(a_N^*)(1-\sigma(a_N^*))\end{pmatrix}\right]$$

P.295より $\begin{cases}(W_N)_{nn} = \sigma(a_n)(1-\sigma(a_n)) \\ W_N \text{ は対角行列}\end{cases}$

$$= \mathrm{Tr}\left[(W_N + C_N^{-1})^{-1}\begin{pmatrix}0 & & & & 0 \\ & \ddots & & & \\ & & \sigma(a_n^*)(1-\sigma(a_n^*))(1-2\sigma(a_n^*)) & & \\ & & & \ddots & \\ 0 & & & & 0\end{pmatrix}\right]$$

$$(W_N + C_N^{-1})^{-1} = \{C_N^{-1}C_N(C_N^{-1} + W_N)\}^{-1} = \{C_N^{-1}(I + C_N W_N)\}^{-1} = (I + C_N W_N)^{-1}C_N$$

$$= \mathrm{Tr}\left[(I + C_N W_N)^{-1}C_N\begin{pmatrix}0 & & & & 0 \\ & \ddots & & & \\ & & \sigma(a_n^*)(1-\sigma(a_n^*))(1-2\sigma(a_n^*)) & & \\ & & & \ddots & \\ 0 & & & & 0\end{pmatrix}\right]$$

$$= \{(I + C_N W_N)^{-1}C_N\}_{nn\text{成分}}\,\sigma(a_n^*)(1-\sigma(a_n^*))(1-2\sigma(a_n^*)) \quad \cdots ③$$

②、③を①に入れて

$$\left\{\frac{\partial}{\partial \mathbf{a}_N^*}\ln p(\mathbf{t}_N|\theta)\right\}^T \frac{\partial \mathbf{a}_N^*}{\partial \theta_j} = -\frac{1}{2}\left(\frac{\partial}{\partial \mathbf{a}_N^*}\ln|W_N + C_N^{-1}|\right)^T \frac{\partial \mathbf{a}_N^*}{\partial \theta_j}$$

$$= -\frac{1}{2}\sum_{n=1}^{N}\frac{\partial}{\partial a_n^*}\ln|W_N + C_N^{-1}|\frac{\partial a_n^*}{\partial \theta_j}$$

$$= -\frac{1}{2}\sum_{n=1}^{N}\{(I + C_N W_N)^{-1}C_N\}_{nn\text{成分}}\,\sigma(a_n^*)(1-\sigma(a_n^*))(1-2\sigma(a_n^*))\frac{\partial a_n^*}{\partial \theta_j} \quad \cdots (6.92)$$

を得る。

(6.94) の導出

(6.84) より

t_N は θ_j に依存しないので

$$\frac{\partial \mathbf{a}_N^*}{\partial \theta_j} = \frac{\partial}{\partial \theta_j} C_N(\mathbf{t}_N - \boldsymbol{\sigma}_N) = \frac{\partial C_N}{\partial \theta_j}(\mathbf{t}_N - \boldsymbol{\sigma}_N) + C_N \frac{\partial}{\partial \theta_j}(\mathbf{t}_N - \boldsymbol{\sigma}_N) = \frac{\partial C_N}{\partial \theta_j}(\mathbf{t}_N - \boldsymbol{\sigma}_N) - C_N \frac{\partial \boldsymbol{\sigma}_N}{\partial \theta_j}$$

ここで

$$\frac{\partial \boldsymbol{\sigma}_N}{\partial \theta_j} = \frac{\partial}{\partial \theta_j}\begin{pmatrix} \sigma(a_1^*) \\ \vdots \\ \sigma(a_N^*) \end{pmatrix} = \begin{pmatrix} \frac{\partial \sigma(a_1^*)}{\partial a_1^*}\frac{\partial a_1^*}{\partial \theta_j} \\ \vdots \\ \frac{\partial \sigma(a_N^*)}{\partial a_N^*}\frac{\partial a_N^*}{\partial \theta_j} \end{pmatrix} = \begin{pmatrix} \frac{\partial \sigma(a_1^*)}{\partial a_1^*} & & 0 \\ & \ddots & \\ 0 & & \frac{\partial \sigma(a_N^*)}{\partial a_N^*} \end{pmatrix}\begin{pmatrix} \frac{\partial a_1^*}{\partial \theta_j} \\ \vdots \\ \frac{\partial a_N^*}{\partial \theta_j} \end{pmatrix} = \begin{pmatrix} \sigma(a_1^*)(1-\sigma(a_1^*)) & & 0 \\ & \ddots & \\ 0 & & \sigma(a_N^*)(1-\sigma(a_N^*)) \end{pmatrix}\begin{pmatrix} \frac{\partial a_1^*}{\partial \theta_j} \\ \vdots \\ \frac{\partial a_N^*}{\partial \theta_j} \end{pmatrix} = W_N \frac{\partial \mathbf{a}_N^*}{\partial \theta_j}$$

なので

$$\frac{\partial \mathbf{a}_N^*}{\partial \theta_j} = \frac{\partial C_N}{\partial \theta_j}(\mathbf{t}_N - \boldsymbol{\sigma}_N) - C_N W_N \frac{\partial \mathbf{a}_N^*}{\partial \theta_j} \quad \cdots (6.93)$$

を得る。$\frac{\partial \mathbf{a}_N^*}{\partial \theta_j}$ について整理すると

$$(I + C_N W_N)\frac{\partial \mathbf{a}_N^*}{\partial \theta_j} = \frac{\partial C_N}{\partial \theta_j}(\mathbf{t}_N - \boldsymbol{\sigma}_N)$$

$$\therefore \frac{\partial \mathbf{a}_N^*}{\partial \theta_j} = (I + C_N W_N)^{-1}\frac{\partial C_N}{\partial \theta_j}(\mathbf{t}_N - \boldsymbol{\sigma}_N) \quad \cdots (6.94)$$

を得る。