3.21

2022 年 5 月 23 日

(問題)

(3.92) はエビデンスの枠組みにおける最適な α の値である。この結果は、次の等式を使って導出することもできる。

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\ln|\mathbf{A}| = \mathrm{Tr}\left(\mathbf{A}^{-1}\frac{\mathrm{d}}{\mathrm{d}\alpha}\mathbf{A}\right) \tag{3.117}$$

実対称行列 $\mathbf{A}$ の固有値展開、および $\mathbf{A}$ の行列式とトレースの固有値表現の標準的結果 (付録 C 参照) を用いて、この等式を証明せよ。そして、(3.117) を用いて (3.86) から (3.92) を導け。

(参考)

$$\mathbf{A} = \alpha\mathbf{I}_M + \beta\mathbf{\Phi}^\top\mathbf{\Phi} \tag{3.81}$$

$$E(\mathbf{m}_N) = \frac{\beta}{2}\left|\left|\mathbf{t} - \mathbf{\Phi}\mathbf{m}_N\right|\right|^2 + \frac{\alpha}{2}\mathbf{m}_N^\top\mathbf{m}_N \tag{3.82}$$

$$\ln p(\mathbf{t}|\alpha,\beta) = \frac{M}{2}\ln\alpha + \frac{N}{2}\ln\beta - E(\mathbf{m}_N) - \frac{1}{2}\ln|\mathbf{A}| - \frac{N}{2}\ln(2\pi) \tag{3.86}$$

$$\gamma = \sum_{i=1}^{M}\frac{\lambda_i}{\lambda_i + \alpha} \tag{3.91}$$

$$\alpha = \frac{\gamma}{\mathbf{m}_N^\top\mathbf{m}_N} \tag{3.92}$$

$$\mathbf{A} = \sum_{i=1}^{M}\lambda_i\mathbf{u}_i\mathbf{u}_i^\top \tag{C.45}$$

$$\mathbf{A}^{-1} = \sum_{i=1}^{M}\frac{1}{\lambda_i}\mathbf{u}_i\mathbf{u}_i^\top \tag{C.46}$$

$$|\mathbf{A}| = \prod_{i=1}^{M}\lambda_i \tag{C.47}$$

(解答)

A の固有値を $\lambda_i (i = 1, \ldots, M)$ とする。また、それぞれの固有値に対応する固有ベクトルを $\mathbf{u_i}(\mathbf{i} = \mathbf{1}, \ldots, \mathbf{M})$ とする。式 (3.81) より、A の固有値はパラメータ α に依存することに注意。なお、$\mathrm{u_i}(\mathrm{i} = 1, \ldots \mathrm{M})$ は正規直行基底であり、

$$\begin{cases} \mathbf{u_i^T u_j} = \delta_{\mathbf{ij}} \\ \sum_{i=1}^{M} \mathbf{u_i u_i^T} = \mathbf{I} \end{cases}$$

となる。式 (3.117) の左辺 $\dfrac{\mathrm{d}}{\mathrm{d}\alpha} \ln |\mathbf{A}|$ を計算する。

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\alpha} \ln |\mathbf{A}| &= \frac{\mathrm{d}}{\mathrm{d}\alpha} \ln \underbrace{\prod_{i=1}^{M} \lambda_i}_{(C.47)} \\ &= \frac{\mathrm{d}}{\mathrm{d}\alpha} \sum_{i=1}^{M} \ln \lambda_i \\ &= \sum_{i=1}^{M} \frac{1}{\lambda_i} \frac{\mathrm{d}\lambda_\mathrm{i}}{\mathrm{d}\alpha} \end{aligned} \tag{1}$$

$$
\begin{aligned}
\mathrm{Tr}\left(\mathbf{A}^{-1}\frac{\mathrm{d}}{\mathrm{d}\alpha}\mathbf{A}\right) &= \mathrm{Tr}\left(\underbrace{\sum_{i=1}^{M}\frac{1}{\lambda_i}\mathbf{u}_i\mathbf{u}_i^\top}_{(C.46)}\frac{\mathrm{d}}{\mathrm{d}\alpha}\sum_{j=1}^{M}\lambda_j\mathbf{u}_j\mathbf{u}_j^\top\right) \\
&= \mathrm{Tr}\left(\sum_{i=1}^{M}\frac{1}{\lambda_i}\mathbf{u}_i\mathbf{u}_i^\top\sum_{j=1}^{M}\left(\frac{\mathrm{d}\lambda_j}{\mathrm{d}\alpha}\mathbf{u}_j\mathbf{u}_j^\top+\lambda_j\frac{\mathrm{d}\mathbf{u}_j}{\mathrm{d}\alpha}\mathbf{u}_j^\top+\lambda_j\mathbf{u}_j\frac{\mathrm{d}\mathbf{u}_j^\top}{\mathrm{d}\alpha}\right)\right) \\
&= \mathrm{Tr}\left(\sum_{i=1}^{M}\frac{1}{\lambda_i}\mathbf{u}_i\mathbf{u}_i^\top\sum_{j=1}^{M}\left(\frac{\mathrm{d}\lambda_j}{\mathrm{d}\alpha}\mathbf{u}_j\mathbf{u}_j^\top+2\lambda_j\mathbf{u}_j\frac{\mathrm{d}\mathbf{u}_j^\top}{\mathrm{d}\alpha}\right)\right) \\
&= \mathrm{Tr}\left(\sum_{i=1}^{M}\frac{1}{\lambda_i}\mathbf{u}_i\mathbf{u}_i^\top\sum_{j=1}^{M}\frac{\mathrm{d}\lambda_j}{\mathrm{d}\alpha}\mathbf{u}_j\mathbf{u}_j^\top+\sum_{i=1}^{M}\frac{1}{\lambda_i}\mathbf{u}_i\mathbf{u}_i^\top\sum_{j=1}^{M}2\lambda_j\mathbf{u}_j\frac{\mathrm{d}\mathbf{u}_j^\top}{\mathrm{d}\alpha}\right) \\
&= \mathrm{Tr}\left(\sum_{i=1}^{M}\sum_{j=1}^{M}\frac{1}{\lambda_i}\frac{\mathrm{d}\lambda_j}{\mathrm{d}\alpha}\mathbf{u}_i\mathbf{u}_i^\top\mathbf{u}_j\mathbf{u}_j^\top+\sum_{i=1}^{M}\sum_{j=1}^{M}\frac{1}{\lambda_i}2\lambda_j\mathbf{u}_i\mathbf{u}_i^\top\mathbf{u}_j\frac{\mathrm{d}\mathbf{u}_j^\top}{\mathrm{d}\alpha}\right) \\
&= \mathrm{Tr}\left(\sum_{i=1}^{M}\frac{1}{\lambda_i}\frac{\mathrm{d}\lambda_i}{\mathrm{d}\alpha}\mathbf{u}_i\mathbf{u}_i^\top+\sum_{i=1}^{M}2\mathbf{u}_i\frac{\mathrm{d}\mathbf{u}_i^\top}{\mathrm{d}\alpha}\right) \\
&= \mathrm{Tr}\left(\sum_{i=1}^{M}\frac{1}{\lambda_i}\frac{\mathrm{d}\lambda_i}{\mathrm{d}\alpha}\mathbf{u}_i\mathbf{u}_i^\top\right)+\mathrm{Tr}\left(\sum_{i=1}^{M}2\mathbf{u}_i\frac{\mathrm{d}\mathbf{u}_i^\top}{\mathrm{d}\alpha}\right) \\
&= \sum_{i=1}^{M}\frac{1}{\lambda_i}\frac{\mathrm{d}\lambda_i}{\mathrm{d}\alpha}+\mathrm{Tr}\left(\sum_{i=1}^{M}\left(\frac{\mathrm{d}\mathbf{u}_i}{\mathrm{d}\alpha}\mathbf{u}_i^\top+\mathbf{u}_i\frac{\mathrm{d}\mathbf{u}_i^\top}{\mathrm{d}\alpha}\right)\right) \\
&= \sum_{i=1}^{M}\frac{1}{\lambda_i}\frac{\mathrm{d}\lambda_i}{\mathrm{d}\alpha}+\mathrm{Tr}\left(\frac{\mathrm{d}}{\mathrm{d}\alpha}\sum_{i=1}^{M}\mathbf{u}_i\mathbf{u}_i^\top\right) \\
&= \sum_{i=1}^{M}\frac{1}{\lambda_i}\frac{\mathrm{d}\lambda_i}{\mathrm{d}\alpha}+\mathrm{Tr}\left(\frac{\mathrm{d}}{\mathrm{d}\alpha}\mathbf{I}\right) \\
&= \sum_{i=1}^{M}\frac{1}{\lambda_i}\frac{\mathrm{d}\lambda_i}{\mathrm{d}\alpha} \qquad (2)
\end{aligned}
$$

式 (1),(2) より、式 (3.117) が示された。

式 (3.86) を α で微分して $=0$ とおく。

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\alpha}\ln p(\mathbf{t}|\alpha,\beta)=0\\
&\Leftrightarrow \frac{M}{2}\frac{\mathrm{d}}{\mathrm{d}\alpha}\ln\alpha+\frac{\mathrm{d}}{\mathrm{d}\alpha}E(\mathbf{m}_N)-\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}\alpha}\ln|\mathbf{A}|=0\\
&\Leftrightarrow \frac{M}{2\alpha}+\frac{\mathrm{d}}{\mathrm{d}\alpha}\left(\frac{\beta}{2}\left|\left|\mathbf{t}-\mathbf{\Phi}\mathbf{m}_N\right|\right|^2+\frac{\alpha}{2}\mathbf{m}_N^\top\mathbf{m}_N\right)-\frac{1}{2}\mathrm{Tr}\left(\mathbf{A}^{-1}\frac{\mathrm{d}}{\mathrm{d}\alpha}\mathbf{A}\right)=0\\
&\Leftrightarrow \frac{M}{2\alpha}+\frac{1}{2}\mathbf{m}_N^\top\mathbf{m}_N-\frac{1}{2}\mathrm{Tr}\left(\mathbf{A}^{-1}\right)=0\ \left(\because \frac{\mathrm{d}}{\mathrm{d}\alpha}\mathbf{A}=\mathbf{I}\right)\\
&\Leftrightarrow \frac{M}{2\alpha}-\frac{1}{2}\mathbf{m}_N^\top\mathbf{m}_N-\frac{1}{2}\sum_{i=1}^{M}\frac{1}{\lambda_i+\alpha}=0\\
&\Leftrightarrow \alpha\mathbf{m}_N^\top\mathbf{m}_N=M-\sum_{i=1}^{M}\frac{\alpha}{\lambda_i+\alpha}\\
&\Leftrightarrow \alpha\mathbf{m}_N^\top\mathbf{m}_N=\sum_{i=1}^{M}\frac{\lambda_i}{\lambda_i+\alpha}\\
&\Leftrightarrow \alpha=\frac{\overbrace{\gamma}^{(3.91)}}{\mathbf{m}_N^\top\mathbf{m}_N}
\end{aligned}
\tag{3}
$$

式 (3) より、式 (3.92) が示された。

(問題 3-23)

対数周辺尤度関数 (3.86) の β に関する最大化が再推定方程式 (3.95) に帰着されることを示すのにすべての段階を、 (3.86) から始めて確かめよ。

(参考)

$$
\ln p(\mathbf{t}|\alpha,\beta)=\frac{M}{2}\ln\alpha+\frac{N}{2}\ln\beta-E(\mathbf{m}_N)-\frac{1}{2}\ln|\mathbf{A}|-\frac{N}{2}\ln(2\pi) \tag{3.86}
$$

$$
\gamma=\sum_{i=1}^{M}\frac{\lambda_i}{\lambda_i+\alpha} \tag{3.91}
$$

$$
\frac{1}{\beta}=\frac{1}{N-\gamma}\sum_{n=1}^{N}\left(t_n-\mathbf{m}_N^\top\boldsymbol{\phi}(\mathbf{x}_n)\right)^2 \tag{3.95}
$$

(解答 3-23)

式 (3.86) の β に関する最大化を考えるため、(3.86) を β に関して偏微分することを考える。

行列 $\beta\mathbf{\Phi}^\top\mathbf{\Phi}$ の固有値を λ_i とする。行列 $\beta\mathbf{\Phi}^\top\mathbf{\Phi}$ を行列 $\mathbf{P}$ で対角化する。

$$
\mathbf{P}^{-1}(\beta\mathbf{\Phi}^\top\mathbf{\Phi})\mathbf{P}=\mathrm{diag}(\lambda_1,\ldots,\lambda_M) \tag{1}
$$

すると、式 (3.81) から、行列 $\mathbf{A}$ にも対角化できる。

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathrm{diag}(\alpha + \lambda_1, \ldots, \alpha + \lambda_M) \tag{2}$$

$\beta\mathbf{\Phi}^\top\mathbf{\Phi}$ の固有値 β に比例するため、$\lambda_i = \beta\aleph_i$ とおいて、β で微分すると

$$\begin{aligned} \frac{d}{d\beta}\lambda_i &= \frac{d}{d\beta}\beta a_i \\ &= a_i \\ &= \frac{\lambda_i}{\beta} \end{aligned} \tag{3}$$

となる。また式 (2) から、$\ln|\mathbf{A}|$ を β で微分すると

$$\begin{aligned} \frac{d}{d\beta}\ln|\mathbf{A}| =& \frac{d}{d\beta}\ln\prod_{i=1}^{M}(\lambda_i + \alpha) \\ =& \frac{d}{d\beta}\sum_{i=1}^{M}\ln(\lambda_i + \alpha) \\ =& \sum_{i=1}^{M}\frac{d}{d(\lambda_i + \alpha)}\ln(\lambda_i + \alpha)\frac{d}{d\beta}(\lambda_i + \alpha) \\ =& \frac{1}{\beta}\sum_{i=1}^{M}\frac{\lambda_i}{\lambda_i + \alpha} \\ =& \frac{\gamma}{\beta} \end{aligned} \tag{4}$$

対数周辺尤度 (3.86) を β で微分して $= 0$ とおく。

$$\begin{aligned} & \frac{\partial}{\partial\beta}\ln p(\mathbf{t}|\alpha, \beta) = 0 \\ & \Leftrightarrow \frac{N}{2\beta} - \frac{1}{2}||\mathbf{t} - \mathbf{\Phi}\mathbf{m}_N||^2 - \frac{\gamma}{2\beta} = 0 \\ & \Leftrightarrow \frac{1}{\beta} = \frac{1}{N - \gamma}||\mathbf{t} - \mathbf{\Phi}\mathbf{m}_N||^2 \\ & \Leftrightarrow \frac{1}{\beta} = \frac{1}{N - \gamma}\sum_{n=1}^{N}\left(t_n - \mathbf{m}_N^\top\boldsymbol{\phi}(\mathbf{x}_n)\right)^2 \end{aligned} \tag{5}$$

式 (5) より、式 (3.95) が示された。