

1 EM アルゴリズム

対数尤度 $\log q(\mathbf{X}; \boldsymbol{\theta})$ の下界は以下のようになり、これを f とおく。

$$f(\mathbf{X}; \boldsymbol{\theta}) = \int q(\mathbf{Z}) \log p(\mathbf{X}, \mathbf{Z}; \boldsymbol{\theta}) d\mathbf{Z} \quad (1)$$

この下界を最大化するパラメータ $\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda}$ を求める。まず、

$$q(\mathbf{Z}) = \prod_{n=1}^N \prod_{k=1}^K \gamma_{nk}^{z_{nk}}$$

$$p(\mathbf{X}, \mathbf{Z}; \boldsymbol{\theta}) = \prod_{n=1}^N \prod_{k=1}^K (\pi_k N(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1}))^{z_{nk}}$$

であり、これを式 (1) に代入して整理すると、次のようになる。

$$\begin{aligned} f(\mathbf{X}; \boldsymbol{\theta}) &= \int \prod_{n=1}^N \prod_{k=1}^K \gamma_{nk}^{z_{nk}} \log \prod_{n=1}^N \prod_{k=1}^K (\pi_k N(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1}))^{z_{nk}} d\mathbf{Z} \\ &= \int \prod_{n=1}^N \prod_{k=1}^K \gamma_{nk}^{z_{nk}} \sum_{n=1}^N \sum_{k=1}^K z_{nk} \log(\pi_k N(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1})) d\mathbf{Z} \\ &= \sum_{n=1}^N \sum_{k=1}^K \int \prod_{n'=1}^N \prod_{k'=1}^K \gamma_{n'k'}^{z_{n'k'}} z_{nk} \log(\pi_k N(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1})) d\mathbf{Z} \\ &= \sum_{n=1}^N \sum_{k=1}^K \gamma_{nk} \log(\pi_k N(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1})) \\ &= \sum_{n=1}^N \sum_{k=1}^K \gamma_{nk} \left\{ \log \pi_k + \frac{1}{2} \log |\boldsymbol{\Lambda}_k| - \frac{1}{2} (\mathbf{x}_n - \boldsymbol{\mu}_k)^T \boldsymbol{\Lambda}_k (\mathbf{x}_n - \boldsymbol{\mu}_k) \right\} + const \end{aligned} \quad (2)$$

まず、最適な $\boldsymbol{\mu}, \boldsymbol{\Lambda}$ を計算するために、式 (2) をそれぞれのパラメータに関して偏微分する。

$$\begin{aligned} \frac{\partial f}{\partial \boldsymbol{\mu}_k} &= -\frac{1}{2} \sum_{n=1}^N \gamma_{nk} \boldsymbol{\Lambda}_k (\mathbf{x}_n - \boldsymbol{\mu}_k) \\ \therefore \boldsymbol{\mu}_k^* &= \frac{\sum_{n=1}^N \gamma_{nk} \mathbf{x}_n}{\sum_{n=1}^N \gamma_{nk}} = \frac{S_k [\mathbf{x}]}{S_k [1]} \\ \frac{\partial f}{\partial \boldsymbol{\Lambda}_k} &= \frac{1}{2} \sum_{n=1}^N \gamma_{nk} \{ \boldsymbol{\Lambda}_k - (\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^T \} \\ \therefore \boldsymbol{\Lambda}_k^* &= \frac{\sum_{n=1}^N \gamma_{nk} (\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^T}{\sum_{n=1}^N \gamma_{nk}} = \frac{S_k [\mathbf{x} \mathbf{x}^T]}{S_k [1]} - \boldsymbol{\mu}_k \boldsymbol{\mu}_k^T \end{aligned}$$

$\boldsymbol{\pi}$ については、 $\sum_{k=1}^K \pi_k = 1$ の条件からラグランジュの未定乗数法を用いる。すなわち、次の式について

各 π_k で微分することを考える.

$$g(\mathbf{X}, \lambda; \boldsymbol{\theta}) = f(\mathbf{X}; \boldsymbol{\theta}) + \lambda \left(\sum_{k=1}^K \pi_k - 1 \right)$$

$$\therefore \frac{\partial g}{\partial \pi_k} = \sum_{n=1}^N \frac{\gamma_{nk}}{\pi_k} + \lambda$$

これをすべての π_k について足し合わせることで, $\lambda = -N$ を得る. よって,

$$\pi_k^* = \frac{-\sum_{n=1}^N \gamma_{nk}}{\lambda} = \frac{S_k[1]}{N}$$

2 変分ベイズ

因子 $q(\boldsymbol{\pi})$, $q(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ の最適解は次のように書けることを確認しておく.

$$\log q^*(\boldsymbol{\pi}) = \log p(\boldsymbol{\pi}) + \langle \log p(\mathbf{Z} | \boldsymbol{\pi}) \rangle_{q(\mathbf{Z})} + \text{const} \quad (3)$$

$$\log q^*(\boldsymbol{\mu}, \boldsymbol{\Lambda}) = \log p(\boldsymbol{\mu}, \boldsymbol{\Lambda}) + \langle \log p(\mathbf{X} | \mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) \rangle_{q(\mathbf{Z})} + \text{const} \quad (4)$$

まず $q^*(\boldsymbol{\pi})$ を求める. 式 (3) 中に現れる確率密度関数を書き下すと, それぞれ次のようになる.

$$p(\boldsymbol{\pi}) = \text{Dir}(\boldsymbol{\pi} | \boldsymbol{\alpha}_0) = \frac{\Gamma(\sum_{k=1}^K \alpha_{0k})}{\prod_{k=1}^K \Gamma(\alpha_{0k})} \prod_{k=1}^K \pi_k^{\alpha_{0k}-1}$$

$$p(\mathbf{Z} | \boldsymbol{\pi}) = \prod_{n=1}^N \prod_{k=1}^K \pi_k^{z_{nk}}$$

また, $q^*(\mathbf{Z}) = \prod_{n=1}^N \prod_{k=1}^K \gamma_{nk}^{z_{nk}}$ より,

$$\langle z_{nk} \rangle_{q(\mathbf{Z})} = \gamma_{nk} \quad (5)$$

これらを利用すると, $\log q^*(\boldsymbol{\pi})$ は次のようになる.

$$\begin{aligned} \log q^*(\boldsymbol{\pi}) &= \sum_{k=1}^K (\alpha_{0k} - 1) \log \pi_k + \sum_{k=1}^K \sum_{n=1}^N \gamma_{nk} \log \pi_k + \text{const} \\ &= \sum_{k=1}^K (\alpha_{0k} + S_k[1] - 1) \log \pi_k + \text{const} \\ &= \sum_{k=1}^K (\alpha_k - 1) \log \pi_k + \text{const} \end{aligned}$$

ただし, $\alpha_k = \alpha_{0k} + S_k[1]$ とする. 最後に両辺の指数をとり, 適当に正規化してやることによって, $q^*(\boldsymbol{\pi})$ はディリクレ分布になる.

$$q^*(\boldsymbol{\pi}) = \text{Dir}(\boldsymbol{\pi} | \boldsymbol{\alpha})$$

次に $q^*(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ を求める．まず，式 (4) 中に現れる確率密度関数を書き下すと，それぞれ次のようになる．

$$p(\boldsymbol{\mu}, \boldsymbol{\Lambda}) = \prod_{k=1}^K N(\boldsymbol{\mu}_k \mid \mathbf{m}_0, (\beta_0 \boldsymbol{\Lambda}_k)^{-1}) W(\boldsymbol{\Lambda}_k \mid \mathbf{W}_0, \nu_0)$$

$$p(\mathbf{X} \mid \mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = \prod_{n=1}^N \prod_{k=1}^K N(\mathbf{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1})^{z_{nk}}$$

これらの密度関数と式 (5) を用いると，式 (4) は次のようになる．

$$\begin{aligned} \log q^*(\boldsymbol{\mu}, \boldsymbol{\Lambda}) &= \sum_{k=1}^K \{ \log N(\boldsymbol{\mu}_k \mid \mathbf{m}_0, (\beta_0 \boldsymbol{\Lambda}_k)^{-1}) + \log W(\boldsymbol{\Lambda}_k \mid \mathbf{W}_0, \nu_0) \} \\ &\quad + \sum_{k=1}^K \sum_{n=1}^N \gamma_{nk} \log N(\mathbf{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1}) + \text{const} \\ &= \sum_{k=1}^K \left[\frac{1}{2} \{ \log |\beta_0 \boldsymbol{\Lambda}_k| - (\boldsymbol{\mu}_k - \mathbf{m}_0)^T (\beta_0 \boldsymbol{\Lambda}_k) (\boldsymbol{\mu}_k - \mathbf{m}_0) \} \right. \\ &\quad + \frac{1}{2} \left\{ \sum_{n=1}^N \gamma_{nk} \log |\boldsymbol{\Lambda}_k| - (\mathbf{x}_n - \boldsymbol{\mu}_k)^T \boldsymbol{\Lambda}_k (\mathbf{x}_n - \boldsymbol{\mu}_k) \right\} \\ &\quad + \left. \left\{ \frac{\nu_0 - D - 1}{2} \log |\boldsymbol{\Lambda}_k| - \frac{1}{2} \text{Tr}(\mathbf{W}_0^{-1} \boldsymbol{\Lambda}_k) \right\} \right] + \text{const} \\ &= \sum_{k=1}^K \left[\frac{1}{2} \underbrace{\{ -(\boldsymbol{\mu}_k - \mathbf{m}_0)^T (\beta_0 \boldsymbol{\Lambda}_k) (\boldsymbol{\mu}_k - \mathbf{m}_0) - (\mathbf{x}_n - \boldsymbol{\mu}_k)^T \boldsymbol{\Lambda}_k (\mathbf{x}_n - \boldsymbol{\mu}_k) \}}_{(a)} \right. \\ &\quad + \underbrace{\left\{ \frac{1}{2} \sum_{n=1}^N \gamma_{nk} \log |\boldsymbol{\Lambda}_k| + \frac{\nu_0 - D - 1}{2} \log |\boldsymbol{\Lambda}_k| - \frac{1}{2} \text{Tr}(\mathbf{W}_0^{-1} \boldsymbol{\Lambda}_k) \right\}}_{(b)} \\ &\quad + \underbrace{\left. \frac{1}{2} \log |\beta_0 \boldsymbol{\Lambda}_k| \right]}_{(c)} + \text{const} \end{aligned} \quad (6)$$

次に，式 (6) 中の (a), (b), (c) について変形していく．

$$\begin{aligned} (a) &= \underbrace{\boldsymbol{\mu}_k^T (\beta_0 + \sum_{n=1}^N \gamma_{nk}) \boldsymbol{\Lambda}_k \boldsymbol{\mu}_k}_{S_k[1]} - 2 \underbrace{\boldsymbol{\mu}_k^T \boldsymbol{\Lambda}_k (\beta_0 \mathbf{m}_0 + \sum_{n=1}^N \gamma_{nk})}_{S_k[1]} + \mathbf{m}_0^T \beta_0 \boldsymbol{\Lambda}_k \mathbf{m}_0 + \sum_{n=1}^N \gamma_{nk} \mathbf{x}_n^T \boldsymbol{\Lambda}_k \mathbf{x}_n \\ &= \underbrace{\boldsymbol{\mu}_k^T (\beta_0 + S_k[1]) \boldsymbol{\Lambda}_k \boldsymbol{\mu}_k}_{\beta_k} - 2 \underbrace{\boldsymbol{\mu}_k^T \boldsymbol{\Lambda}_k (\beta_0 + S_k[1])}_{\beta_k} \underbrace{\frac{\beta_0 \mathbf{m}_0 + S_k[\mathbf{x}]}{\beta_0 + S_k[1]}}_{\mathbf{m}_k} + \mathbf{m}_0^T \beta_0 \boldsymbol{\Lambda}_k \mathbf{m}_0 + \sum_{n=1}^N \gamma_{nk} \mathbf{x}_n^T \boldsymbol{\Lambda}_k \mathbf{x}_n \\ &= (\boldsymbol{\mu}_k - \mathbf{m}_k)^T \beta_k \boldsymbol{\Lambda}_k (\boldsymbol{\mu}_k - \mathbf{m}_k) - \mathbf{m}_k^T \beta_k \boldsymbol{\Lambda}_k \mathbf{m}_k + \mathbf{m}_0^T \beta_0 \boldsymbol{\Lambda}_k \mathbf{m}_0 + \sum_{n=1}^N \gamma_{nk} \mathbf{x}_n^T \boldsymbol{\Lambda}_k \mathbf{x}_n \end{aligned} \quad (7)$$

$$\begin{aligned}
(b) &= \frac{1}{2} S_k[1] \log |\mathbf{\Lambda}_k| + \frac{\nu_0 - D - 1}{2} \log |\mathbf{\Lambda}_k| \\
&= \frac{\nu_0 + S_k[1] - D - 1}{2} \log |\mathbf{\Lambda}_k| \\
&= \frac{\nu_k - D - 1}{2} \log |\mathbf{\Lambda}_k|
\end{aligned} \tag{8}$$

$$(c) = \frac{1}{2} \log |\mathbf{\Lambda}_k| + \text{const} \tag{9}$$

ここで,

$$\beta_k = \beta_0 + S_k[1], \quad \mathbf{m}_k = \frac{\beta_0 \mathbf{m}_0 + S_k[\mathbf{x}]}{\beta_0 + S_k[1]}, \quad \nu_k = \nu_0 + S_k[1]$$

とおいた. 式 (7),(8),(9) を式 (6) に戻して整理すると,

$$\begin{aligned}
\log q^*(\boldsymbol{\mu}, \mathbf{\Lambda}) &= \sum_{k=1}^K \left[\left\{ \frac{1}{2} \log |\mathbf{\Lambda}_k| - \frac{1}{2} (\boldsymbol{\mu}_k - \mathbf{m}_k)^T \beta_k \mathbf{\Lambda}_k (\boldsymbol{\mu}_k - \mathbf{m}_k) \right\} \right. \\
&\quad \left. - \frac{1}{2} \underbrace{\left\{ Tr(\mathbf{W}_0^{-1} \mathbf{\Lambda}_k) - \mathbf{m}_k^T \beta_k \mathbf{\Lambda}_k \mathbf{m}_k + \mathbf{m}_0^T \beta_0 \mathbf{\Lambda}_k \mathbf{m}_0 + \sum_{n=1}^N \gamma_{nk} \mathbf{x}_n^T \mathbf{\Lambda}_k \mathbf{x}_n \right\}}_{(d)} \right] \\
&\quad \left. + \frac{\nu_k - D - 1}{2} \log |\mathbf{\Lambda}_k| \right] + \text{const}
\end{aligned} \tag{10}$$

次に, 式 (10) 中の (d) について変形する.

$$\begin{aligned}
(d) &= Tr(\mathbf{W}_0^{-1} \mathbf{\Lambda}_k) - Tr(\mathbf{m}_k^T \beta_k \mathbf{\Lambda}_k \mathbf{m}_k) + Tr(\mathbf{m}_0^T \beta_0 \mathbf{\Lambda}_k \mathbf{m}_0) + Tr\left(\sum_{n=1}^N \gamma_{nk} \mathbf{x}_n^T \mathbf{\Lambda}_k \mathbf{x}_n\right) \\
&= Tr(\mathbf{W}_0^{-1} \mathbf{\Lambda}_k) - Tr(\mathbf{m}_k \mathbf{m}_k^T \beta_k \mathbf{\Lambda}_k) + Tr(\mathbf{m}_0 \mathbf{m}_0^T \beta_0 \mathbf{\Lambda}_k) + Tr\left(\underbrace{\sum_{n=1}^N \gamma_{nk} \mathbf{x}_n \mathbf{x}_n^T \mathbf{\Lambda}_k}_{S_k[\mathbf{x} \mathbf{x}^T]}\right) \\
&= Tr((\mathbf{W}_0^{-1} - \mathbf{m}_k \mathbf{m}_k^T \beta_k + \mathbf{m}_0 \mathbf{m}_0^T \beta_0 + S_k[\mathbf{x} \mathbf{x}^T]) \mathbf{\Lambda}_k) \\
&= Tr(\mathbf{W}_k^{-1} \mathbf{\Lambda}_k)
\end{aligned} \tag{11}$$

ここで, $\mathbf{W}_k^{-1} = \mathbf{W}_0^{-1} - \mathbf{m}_k \mathbf{m}_k^T \beta_k + \mathbf{m}_0 \mathbf{m}_0^T \beta_0 + S_k[\mathbf{x} \mathbf{x}^T]$ とした. 最後に, 式 (11) を (10) に戻して,

$$\begin{aligned}
\log q^*(\boldsymbol{\mu}, \mathbf{\Lambda}) &= \sum_{k=1}^K \left[\left\{ \frac{1}{2} \log |\mathbf{\Lambda}_k| - \frac{1}{2} (\boldsymbol{\mu}_k - \mathbf{m}_k)^T \beta_k \mathbf{\Lambda}_k (\boldsymbol{\mu}_k - \mathbf{m}_k) \right\} \right. \\
&\quad \left. + \left\{ \frac{\nu_k - D - 1}{2} \log |\mathbf{\Lambda}_k| - \frac{1}{2} Tr(\mathbf{W}_k^{-1} \mathbf{\Lambda}_k) \right\} \right] + \text{const} \\
\therefore q^*(\boldsymbol{\mu}, \mathbf{\Lambda}) &= \prod_{k=1}^K N(\boldsymbol{\mu}_k \mid \mathbf{m}_k, (\beta_0 \mathbf{\Lambda}^{-1})) W(\mathbf{\Lambda}_k \mid \mathbf{W}_k, \nu_k)
\end{aligned} \tag{12}$$