第8章 混合ガウスモデルの最尤推定

8.3 EM アルゴリズム

07t4072f　　岡崎　駿

EMアルゴリズムとは

・EMアルゴリズムとは

→勾配法におけるステップ幅の決定問題を回避できる最適化手法

→入力ベクトル**x**が部分的にしか観測できない場合(不完全データ)に対する場合の最尤推定量のアルゴリズム

→ Expectation-step(期待値)とMaximization-step(最大化)の二つのステップを繰り返すことにより、確率パラメータを更新していく

EMアルゴリズムの計算方法(1)

1. 解の初期値 $\{\hat{\omega}_\ell, \hat{\boldsymbol{\mu}}_\ell, \hat{\sigma}_\ell\}_{\ell=1}^m$ を適当に定める。
2. Eステップ：現在の解 $\{\hat{\omega}_\ell, \hat{\boldsymbol{\mu}}_\ell, \hat{\sigma}_\ell\}_{\ell=1}^m$ から媒介変数 $\{\hat{\eta}_{i,\ell}\}_{i=1}^n{}_{\ell=1}^m$ を計算する。

$$\boxed{\hat{\eta}_{i,\ell} \leftarrow \frac{\hat{\omega}_\ell \phi(\mathbf{x}_i; \hat{\boldsymbol{\mu}}_\ell, \hat{\sigma}_\ell^2)}{\sum_{\ell'=1}^m \hat{\omega}_{\ell'} \phi(\mathbf{x}_i; \hat{\boldsymbol{\mu}}_{\ell'}, \hat{\sigma}_{\ell'}^2)}}$$

EMアルゴリズムの計算方法(2)

3.Mステップ:現在の媒介変数$\{\hat{\eta}_{i,\ell}\}_{i=1}^{n}{}_{\ell=1}^{m}$から、
解$\{\hat{\omega}_\ell, \hat{\boldsymbol{\mu}}_\ell, \hat{\sigma}_\ell\}_{\ell=1}^{m}$を計算する。

$$\hat{\omega}_\ell \leftarrow \frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_{i,\ell}$$

$$\hat{\boldsymbol{\mu}}_\ell \leftarrow \frac{\sum_{i=1}^{n}\hat{\eta}_{i,\ell}\mathbf{x}_i}{\sum_{i'=1}^{n}\hat{\eta}_{i',\ell}}$$

$$\hat{\sigma}_\ell \leftarrow \sqrt{\frac{\sum_{i=1}^{n}\hat{\eta}_{i,\ell}(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_\ell)^{\mathrm{T}}(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_\ell)}{d\sum_{i'=1}^{n}\hat{\eta}_{i',\ell}}}$$

4.2～3を収束するまで繰り返す。

EMアルゴリズムの意味(1)

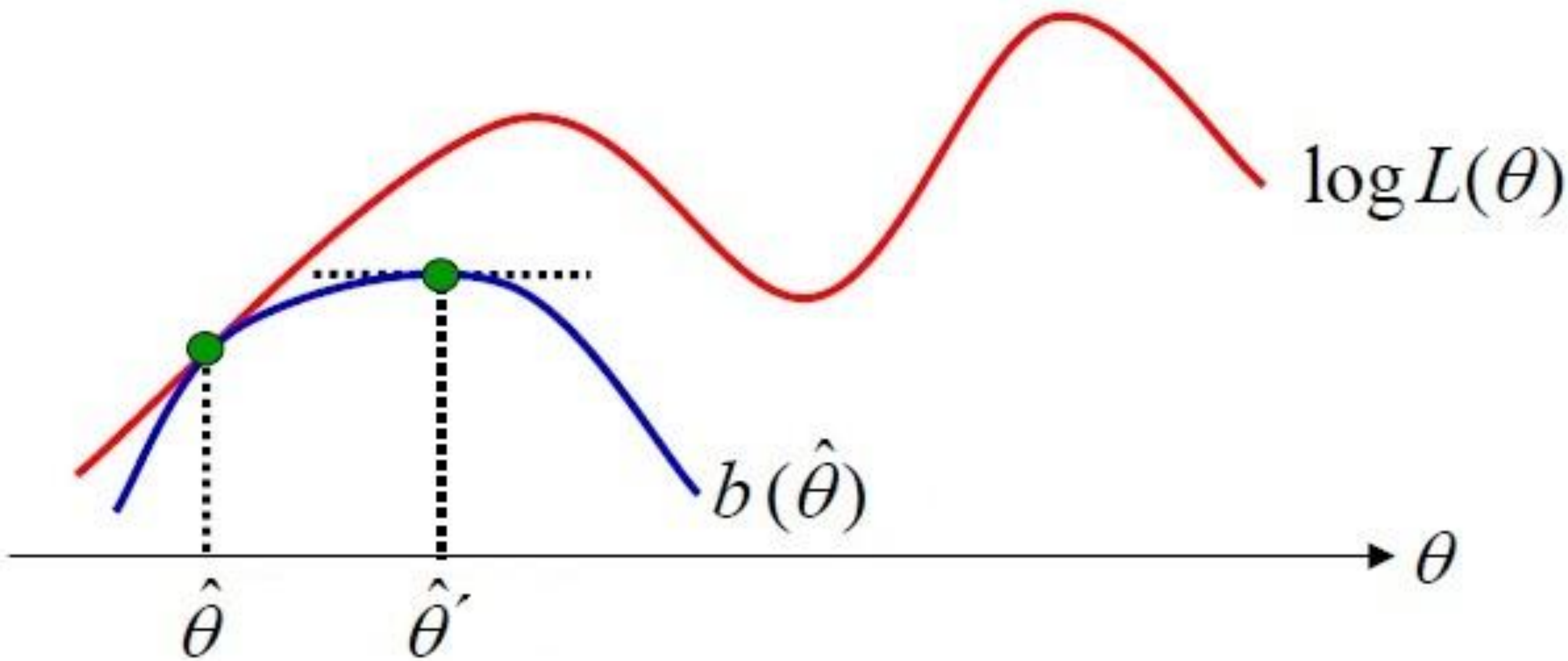
- Eステップ：現在の解 $\hat{\boldsymbol{\theta}}$ を通る対数尤度 $\log L(\boldsymbol{\theta})$ の下界 $b(\boldsymbol{\theta})$ を求めることに対応。すなわち、任意の $\boldsymbol{\theta}$ に対して

$$\log L(\boldsymbol{\theta}) \geq b(\boldsymbol{\theta}), \log L(\hat{\boldsymbol{\theta}}) = b(\hat{\boldsymbol{\theta}})$$

を満たす下界 $b(\boldsymbol{\theta})$ を求めることに対応

- Mステップ：下界 $b(\boldsymbol{\theta})$ を最大にする値 $\hat{\boldsymbol{\theta}}'$ を求めることに対応

EMアルゴリズムの意味(2)



EMアルゴリズムの意味(2)

- EMアルゴリズムによって対数尤度が減少することはない
- Eステップにおける下界の導出が重要

ジェンセンの不等式(1)

非負かつ和が1の変数$\{\eta_\ell\}_{\ell=1}^m$

$$\eta_1,\ldots,\eta_m \geq 0, \sum_{\ell=1}^m \eta_\ell = 1$$

に対して

$$\log\left(\sum_{\ell=1}^m \eta_\ell u_\ell\right) \geq \sum_{\ell=1}^m \eta_\ell \log u_\ell$$

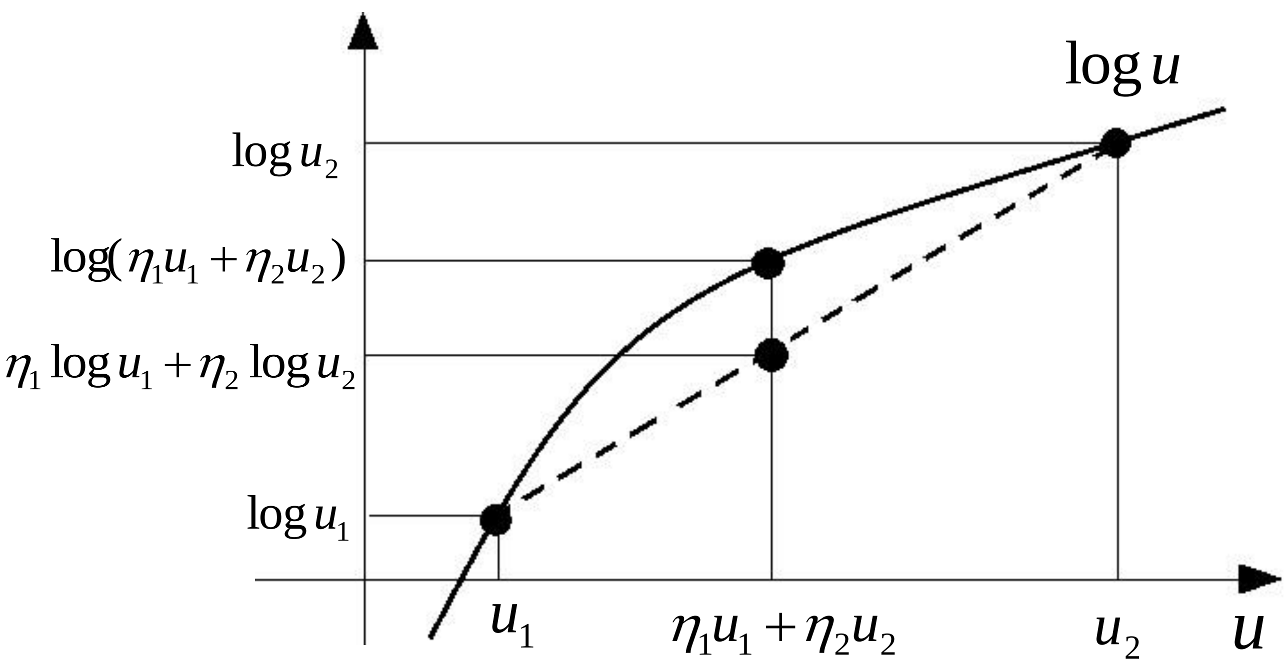
が与えられる。

ジェンセンの不等式(2)

- m=2に対するジェンセンの不等式（例）

$$\log(\eta_1 u_1 + \eta_2 u_2) \geq \eta_1 \log u_1 + \eta_2 \log u_2$$

- 解釈(対数関数は常に凸)



Eステップの下界の導出(1)

$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log(\sum_{\ell=1}^{m} \omega_\ell \phi(\mathbf{x}_i; \boldsymbol{\mu}_\ell, \sigma_\ell^2))$$

$$= \sum_{i=1}^{n} \log(\sum_{\ell=1}^{m} \hat{\eta}_{i,\ell} \frac{\omega_\ell \phi(\mathbf{x}_i; \boldsymbol{\mu}_\ell, \sigma_\ell^2)}{\hat{\eta}_{i,\ell}})$$

ただし、

$$\hat{\eta}_{i,\ell} := \frac{\hat{\omega}_\ell \phi(\mathbf{x}_i; \hat{\boldsymbol{\mu}}_\ell, \hat{\sigma}_\ell^2)}{\sum_{\ell'=1}^{m} \hat{\omega}_{\ell'} \phi(\mathbf{x}_i; \hat{\boldsymbol{\mu}}_{\ell'}, \hat{\sigma}_{\ell'}^2)}$$

ジェンセンの不等式を適応

$$\log L(\boldsymbol{\theta}) \geq \sum_{i=1}^{n} \sum_{\ell=1}^{m} \hat{\eta}_{i,\ell} \log(\frac{\omega_\ell \phi(\mathbf{x}_i; \boldsymbol{\mu}_\ell, \sigma_\ell^2)}{\hat{\eta}_{i,\ell}}) =: b(\boldsymbol{\theta})$$

Eステップの下界の導出(2)

$$\hat{\eta}_{i,\ell} := \frac{\hat{\omega}_\ell \phi(\mathbf{x}_i ; \hat{\boldsymbol{\mu}}_\ell, \hat{\sigma}_\ell^2)}{\sum_{\ell'=1}^{m} \hat{\omega}_{\ell'} \phi(\mathbf{x}_i ; \hat{\boldsymbol{\mu}}_{\ell'}, \hat{\sigma}_{\ell'}^2)}$$ より、

$$\frac{\hat{\omega}_\ell \phi(\mathbf{x}_i ; \hat{\boldsymbol{\mu}}_\ell, \hat{\sigma}_\ell^2)}{\hat{\eta}_{i,\ell}} = \sum_{\ell'=1}^{m} \hat{\omega}_{\ell'} \phi(\mathbf{x}_i ; \hat{\boldsymbol{\mu}}_{\ell'}, \hat{\sigma}_{\ell'}^2), \quad \sum_{\ell=1}^{m} \hat{\eta}_{i,\ell} = 1$$

が成り立つ。これより先ほど導出した$b(\boldsymbol{\theta})$の式を変形させる。

$$\begin{aligned}
b(\hat{\boldsymbol{\theta}}) &= \sum_{i=1}^{n} \sum_{\ell=1}^{m} \hat{\eta}_{i,\ell} \log\left(\frac{\hat{\omega}_\ell \phi(\mathbf{x}_i ; \hat{\boldsymbol{\mu}}_\ell, \hat{\sigma}_\ell^2)}{\hat{\eta}_{i,\ell}} \right) \\
&= \sum_{i=1}^{n} \sum_{\ell=1}^{m} \hat{\eta}_{i,\ell} \log\left(\sum_{\ell'=1}^{m} \hat{\omega}_{\ell'} \phi(\mathbf{x}_i ; \hat{\boldsymbol{\mu}}_{\ell'}, \hat{\sigma}_{\ell'}^2) \right) \\
&= \sum_{i=1}^{n} \left(\sum_{\ell=1}^{m} \hat{\eta}_{i,\ell} \right) \log\left(\sum_{\ell'=1}^{m} \hat{\omega}_{\ell'} \phi(\mathbf{x}_i ; \hat{\boldsymbol{\mu}}_{\ell'}, \hat{\sigma}_{\ell'}^2) \right) \\
&= \sum_{i=1}^{n} \log\left(\sum_{\ell'=1}^{m} \hat{\omega}_{\ell'} \phi(\mathbf{x}_i ; \hat{\boldsymbol{\mu}}_{\ell'}, \hat{\sigma}_{\ell'}^2) \right) \\
&= \log L(\hat{\boldsymbol{\theta}})
\end{aligned}$$

Mステップでのパラメータ導出

- 下界を$b(\boldsymbol{\theta})$ 最大にするパラメータ $\hat{\boldsymbol{\theta}}'$ を求める。
  - $\hat{\boldsymbol{\theta}}'$ は以下を満たす

$$\left\{\begin{array}{l} \left.\dfrac{\partial}{\partial \gamma_\ell} b(\boldsymbol{\theta})\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}'} = 0 \\ \left.\dfrac{\partial}{\partial \boldsymbol{\mu}_\ell} b(\boldsymbol{\theta})\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}'} = \mathbf{0}_d \\ \left.\dfrac{\partial}{\partial \sigma_\ell} b(\boldsymbol{\theta})\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}'} = 0 \end{array}\right. \Rightarrow \left\{\begin{array}{l} \hat{\omega}'_\ell = \dfrac{1}{n}\sum_{i=1}^{n} \hat{\eta}_{i,\ell} \\ \hat{\boldsymbol{\mu}}'_\ell = \dfrac{\sum_{i=1}^{n} \hat{\eta}_{i,\ell}\mathbf{x}_i}{\sum_{i'=1}^{n} \hat{\eta}_{i',\ell}} \\ \hat{\sigma}'_\ell = \sqrt{\dfrac{\sum_{i=1}^{n} \hat{\eta}_{i,\ell}(\mathbf{x}_i - \hat{\boldsymbol{\mu}}'_\ell)^{\mathrm{T}}(\mathbf{x}_i - \hat{\boldsymbol{\mu}}'_\ell)}{d\sum_{i'=1}^{n} \hat{\eta}_{i',\ell}}} \end{array}\right.$$

結論

- EステップとMステップの繰り返しによって対数尤度は減少しない。

- EMアルゴリズムは適当な条件のもとで局所的最適解に収束する。