Mixture Models & The EM Algorithm

Toru Wakahara

[1] 確率密度関数の推定

- セミパラメトリックな方法

[2] 混合モデル

- 基本的な考え方
- 最尤法
- The EM algorithm
- The EM algorithm の適用例

確率密度関数推定の方法

[1] パラメトリックな方法

確率密度関数として特定の関数形を選択し，関数パラメータをデータセットに合うように最適化する.

[2] ノンパラメトリックな方法

自由度の大きな汎用関数を用いて，データセットを忠実に反映するように確率密度関数を構成する.

[3] セミパラメトリックな方法

特定の関数形の重ね合せ（混合モデル）でデータセットの分布を近似する.

セミパラメトリックな方法

特定の関数形の重ね合せにより
データを忠実に表現　→　混合モデル

(1) 最尤法
正規分布の混合モデルで，最尤推定により分布パラメータや混合比を決定する．
(2) The EM algorithm
学習データを用いて，expectation step → maximization stepを反復して，混合モデルの反復解法を与える．

混合モデルの基本的考え方(1)

カーネル法: N 点での核関数の重ね合せ

$$\tilde{p}(\boldsymbol{x}) = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{(2\pi h^2)^{d/2}}\exp\left\{-\frac{\|\boldsymbol{x}-\boldsymbol{x}_n\|^2}{2h^2}\right\}$$

混合モデル: M 個の分布の重ね合せ

$$p(\boldsymbol{x}) = \sum_{j=1}^{M} p(\boldsymbol{x}\mid j)P(j)$$

$$\sum_{j=1}^{M} P(j) = 1,\ \ 0 \le P(j) \le 1,\quad \int p(\boldsymbol{x}\mid j)\,d\boldsymbol{x} = 1$$

混合モデルの基本的考え方(2)

混合モデル: M 個の分布の重ね合せ

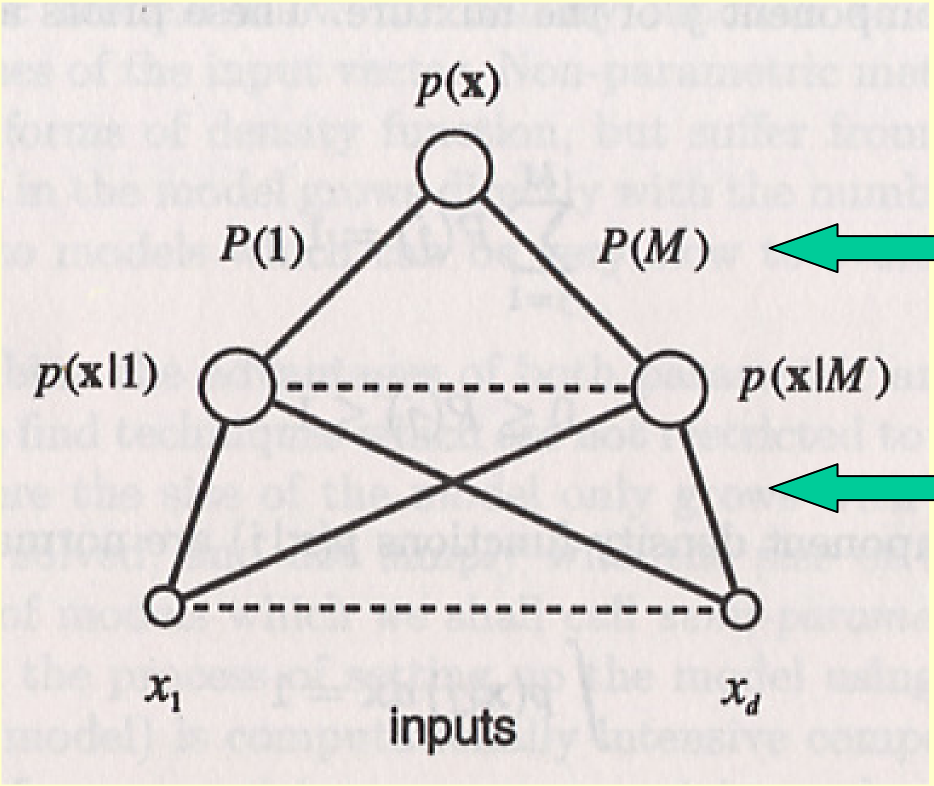
$$p(\boldsymbol{x}) = \sum_{j=1}^{M} p(\boldsymbol{x} \mid j) P(j)$$

$P(j)$: j 番目の分布の事前確率

$p(\boldsymbol{x} \mid j)$: j 番目の分布の確率密度関数

$$p(\boldsymbol{x} \mid j) = \frac{1}{(2\pi\sigma_j^2)^{d/2}} \exp\left\{ -\frac{\left\| \boldsymbol{x} - \boldsymbol{\mu}_j \right\|^2}{2\sigma_j^2} \right\}$$

混合モデルの基本的考え方(3)



$P(j)$:混合パラメータ

$\boldsymbol{\mu}_j, \sigma_j$:分布パラメータ

混合モデル: $P(j)$, μ_j, σ_j をデータセットを用いてどう決定するか

最尤法の適用

「尤度＝データセット X の生起確率」の最大化

⇒ マイナス対数尤度 E の最小化

$$E = -\ln L = -\ln \prod_{n=1}^{N} p(\boldsymbol{x}_n) = -\sum_{n=1}^{N} \ln p(\boldsymbol{x}_n)$$

$$= -\sum_{n=1}^{N} \ln \left\{ \sum_{j=1}^{M} p(\boldsymbol{x}_n \mid j) P(j) \right\}$$

$$\rightarrow \quad \min \ \text{for} \ \boldsymbol{\mu}_j, \sigma_j, P(j)$$

μ_jの最尤推定

$$\frac{\partial E}{\partial \boldsymbol{\mu}_j} = -\sum_{n=1}^{N} \frac{P(j)}{p(\boldsymbol{x}_n)} \frac{\partial p(\boldsymbol{x}_n \mid j)}{\partial \boldsymbol{\mu}_j}$$

$$= \sum_{n=1}^{N} \frac{P(j)\, p(\boldsymbol{x}_n \mid j)}{p(\boldsymbol{x}_n)} \frac{(\boldsymbol{\mu}_j - \boldsymbol{x}_n)}{\sigma_j^2}$$

$$= \sum_{n=1}^{N} P(j \mid \boldsymbol{x}_n) \frac{(\boldsymbol{\mu}_j - \boldsymbol{x}_n)}{\sigma_j^2} = \boldsymbol{0}$$

$$\therefore \hat{\boldsymbol{\mu}}_j = \frac{\sum_n P(j \mid \boldsymbol{x}_n)\, \boldsymbol{x}_n}{\sum_n P(j \mid \boldsymbol{x}_n)}$$

σ_jの最尤推定

$$\frac{\partial E}{\partial \sigma_j} = -\sum_{n=1}^{N} \frac{P(j)}{p(\boldsymbol{x}_n)} \frac{\partial p(\boldsymbol{x}_n \mid j)}{\partial \sigma_j}$$

$$= \sum_{n=1}^{N} P(j \mid \boldsymbol{x}_n) \left\{ \frac{d}{\sigma_j} - \frac{\| \boldsymbol{x}_n - \boldsymbol{\mu}_j \|^2}{\sigma_j^3} \right\} = 0$$

$$\therefore \hat{\sigma}_j^2 = \frac{1}{d} \frac{\sum_n P(j \mid \boldsymbol{x}_n) \| \boldsymbol{x}_n - \hat{\boldsymbol{\mu}}_j \|^2}{\sum_n P(j \mid \boldsymbol{x}_n)}$$

$P(j)$ の最尤推定

$$P(j)=\frac{\exp(\gamma_j)}{\sum_{k=1}^{M}\exp(\gamma_k)} \quad \left(0\le P(j)\le 1,\ \sum_{j=1}^{M}P(j)=1\right)$$

と $P(j)$ を the softmax function で巧妙に表現すると

$$\frac{\partial E}{\partial \gamma_j}=\sum_{k=1}^{M}\frac{\partial E}{\partial P(k)}\frac{\partial P(k)}{\partial \gamma_j}=-\sum_{n=1}^{N}\left\{P(j\,|\,\boldsymbol{x}_n)-P(j)\right\}=0$$

$$\therefore \hat{P}(j)=\frac{1}{N}\sum_{n=1}^{N}P(j\,|\,\boldsymbol{x}_n)$$

最尤法の解

$$\hat{\boldsymbol{\mu}}_j = \frac{\sum_n P(j \mid \boldsymbol{x}_n)\, \boldsymbol{x}_n}{\sum_n P(j \mid \boldsymbol{x}_n)} \qquad \hat{\sigma}_j^2 = \frac{1}{d} \frac{\sum_n P(j \mid \boldsymbol{x}_n) \| \boldsymbol{x}_n - \hat{\boldsymbol{\mu}}_j \|^2}{\sum_n P(j \mid \boldsymbol{x}_n)}$$

$$\hat{P}(j) = \frac{1}{N} \sum_{n=1}^{N} P(j \mid \boldsymbol{x}_n) \qquad P(j \mid \boldsymbol{x}_n) = \frac{p(\boldsymbol{x}_n \mid j)\, P(j)}{p(\boldsymbol{x}_n)}$$

最尤解を求めるには $P(j \mid \boldsymbol{x}_n)$ が必要である.
しかし, $P(j \mid \boldsymbol{x}_n)$ はベイズの定理により
$p(\boldsymbol{x}_n \mid j),\ P(j)$ が分からないと求まらない！

最尤法（E 最小化）の反復解法へ

パラメータの更新による尤度の変化に注目

$$E^{\text{new}} - E^{\text{old}} = -\sum_{n=1}^{N} \ln p^{\text{new}}(\boldsymbol{x}_n) + \sum_{n=1}^{N} \ln p^{\text{old}}(\boldsymbol{x}_n)$$

$$= -\sum_n \ln \left\{ \frac{p^{\text{new}}(\boldsymbol{x}_n)}{p^{\text{old}}(\boldsymbol{x}_n)} \right\}$$

$$= -\sum_n \ln \left\{ \frac{\sum_j P^{\text{new}}(j)\, p^{\text{new}}(\boldsymbol{x}_n \mid j)}{p^{\text{old}}(\boldsymbol{x}_n)} \frac{P^{\text{old}}(j \mid \boldsymbol{x}_n)}{P^{\text{old}}(j \mid \boldsymbol{x}_n)} \right\}$$

The EM algorithm (1)

$$E^{\mathrm{new}} - E^{\mathrm{old}} = -\sum_n \ln\left\{ \frac{\sum_j P^{\mathrm{new}}(j)\, p^{\mathrm{new}}(\boldsymbol{x}_n \mid j)}{p^{\mathrm{old}}(\boldsymbol{x}_n)} \frac{P^{\mathrm{old}}(j \mid \boldsymbol{x}_n)}{P^{\mathrm{old}}(j \mid \boldsymbol{x}_n)} \right\}$$

上式で Jensen 不等式を用いると

$$\lambda_j \geq 0, \sum_j \lambda_j = 1 \rightarrow \ln\left(\sum_j \lambda_j x_j \right) \geq \sum_j \lambda_j \ln x_j$$

$$E^{\mathrm{new}} - E^{\mathrm{old}} \leq -\sum_n \sum_j P^{\mathrm{old}}(j \mid \boldsymbol{x}_n) \ln\left\{ \frac{P^{\mathrm{new}}(j)\, p^{\mathrm{new}}(\boldsymbol{x}_n \mid j)}{p^{\mathrm{old}}(\boldsymbol{x}_n)\, P^{\mathrm{old}}(j \mid \boldsymbol{x}_n)} \right\} \equiv Q$$

Q の最小化 → E^{new} の上限の最小化

The EM algorithm (2)

Expectation step: E^{new} の上限値を評価

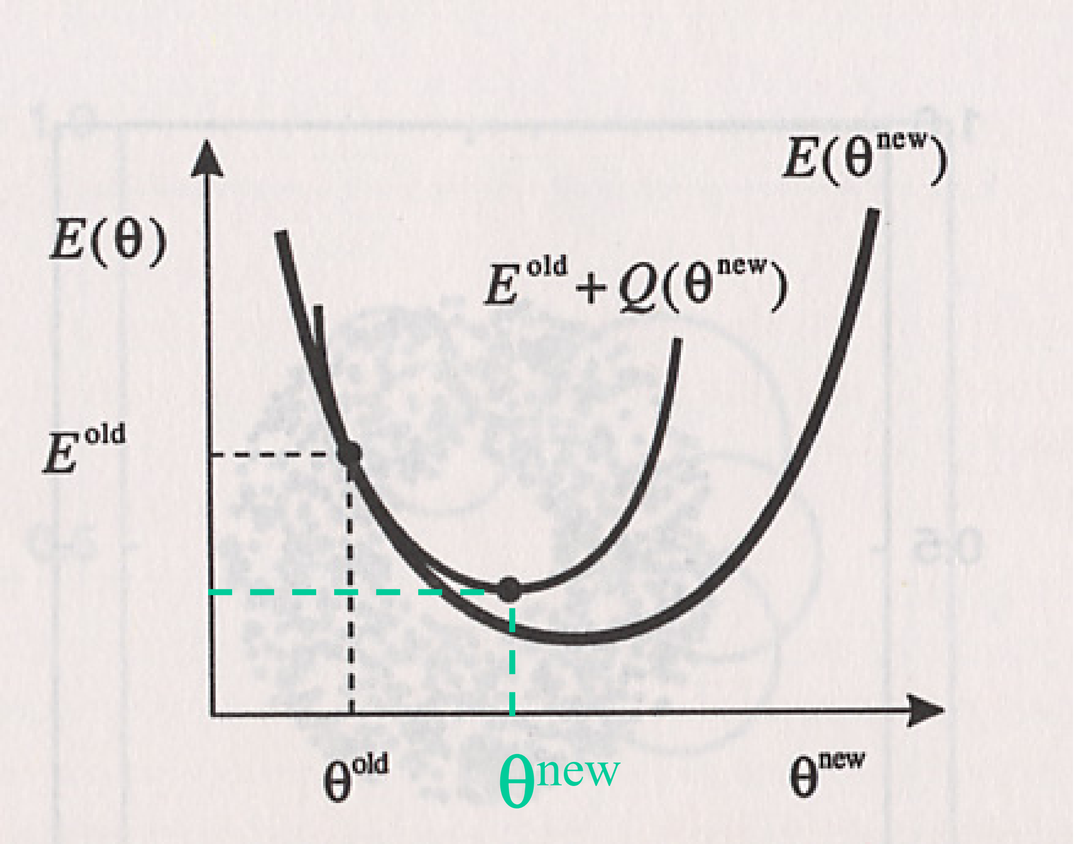
$$\tilde{Q} = -\sum_n \sum_j P^{\mathrm{old}}(j \mid \boldsymbol{x}_n) \ln \left\{ P^{\mathrm{new}}(j)\, p^{\mathrm{new}}(\boldsymbol{x}_n \mid j) \right\}$$

反復解法

Maximization step: E^{new} の上限値の最小化

$$\tilde{Q} + \lambda \left(\sum_j P^{\mathrm{new}}(j) - 1 \right) \quad \rightarrow \quad \min \ \text{for} \ \theta^{\mathrm{new}}$$

The EM algorithm (3)



$$E^{new} \leq E^{old} + Q(\theta^{new})$$

$Q(\theta^{new})$ の最小化 → E^{new} の上限の最小化

The EM algorithm (4)

正規分布の場合:

$$\tilde{Q} = -\sum_n \sum_j P^{\mathrm{old}}(j \mid \boldsymbol{x}_n) \times \left\{ \ln P^{\mathrm{new}}(j) - d \ln \sigma_j^{\mathrm{new}} - \frac{\| \boldsymbol{x}_n - \boldsymbol{\mu}_j^{\mathrm{new}} \|^2}{2(\sigma_j^{\mathrm{new}})^2} \right\} + \mathrm{const.}$$

Lagrange の未定乗数法

$$\tilde{Q} + \lambda \left(\sum_j P^{\mathrm{new}}(j) - 1 \right) \quad \rightarrow \quad \min \ \text{for} \ \boldsymbol{\mu}_j^{\mathrm{new}}, \sigma_j^{\mathrm{new}}, P^{\mathrm{new}}(j)$$

The EM algorithm (5)

正規分布の混合モデルでの EM 解:

$$\boldsymbol{\mu}_j^{\text{new}} = \frac{\sum_n P^{\text{old}}(j \mid \boldsymbol{x}_n)\,\boldsymbol{x}_n}{\sum_n P^{\text{old}}(j \mid \boldsymbol{x}_n)} \qquad P^{\text{new}}(j) = \frac{1}{N}\sum_{n=1}^{N} P^{\text{old}}(j \mid \boldsymbol{x}_n)$$

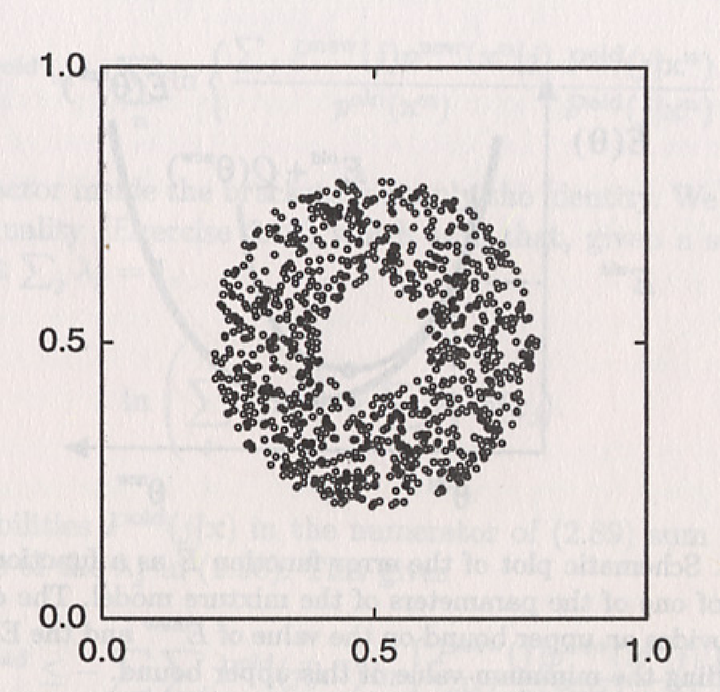
$$(\sigma_j^{\text{new}})^2 = \frac{1}{d}\,\frac{\sum_n P^{\text{old}}(j \mid \boldsymbol{x}_n)\,\|\boldsymbol{x}_n - \boldsymbol{\mu}_j^{\text{new}}\|^2}{\sum_n P^{\text{old}}(j \mid \boldsymbol{x}_n)}$$

$$\text{where} \quad P^{\text{old}}(j \mid \boldsymbol{x}_n) = \frac{p^{\text{old}}(\boldsymbol{x}_n \mid j)\,P^{\text{old}}(j)}{\sum_k p^{\text{old}}(\boldsymbol{x}_n \mid k)\,P^{\text{old}}(k)}$$

課題3

混合モデルのEM解を体験してみよう.

具体例として，下図のように，中心 (0.5, 0.5) の環状（内径：0.1，外径：0.3）内に分布するデータ1000点を一様乱数を用いて発生しなさい.



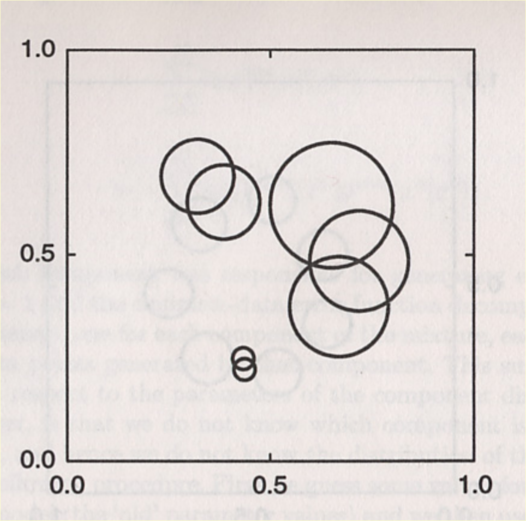
この分布を正規分布の混合モデルで近似してみる.

課題3（続き）

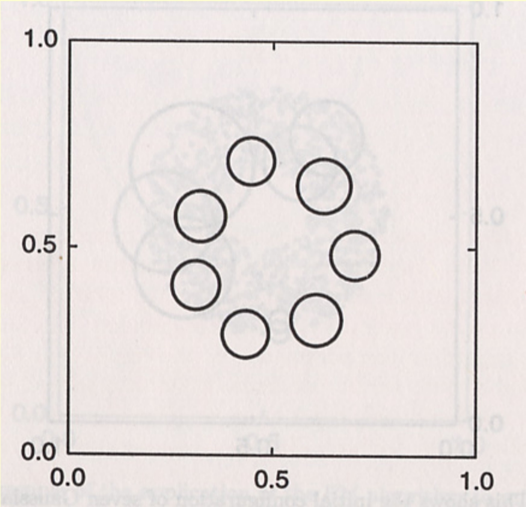
モデルセンタはランダムに M 点選択して,

$P(j)=1/M$,　$\sigma_j=$ 最近傍センタまでの距離,

で初期化し, EM 解を求めてみなさい.



正規分布の
初期配置（$M=7$）



正規分布の最終配置
（反復20回）