EMアルゴリズム

(Expectation –Maximization)

渡辺澄夫

目標

$$L(w) = \sum_{i=1}^{n} \log p(x_i|w)$$ を最大にする w を求めたい

しかし、例えば、混合正規分布では解けない。

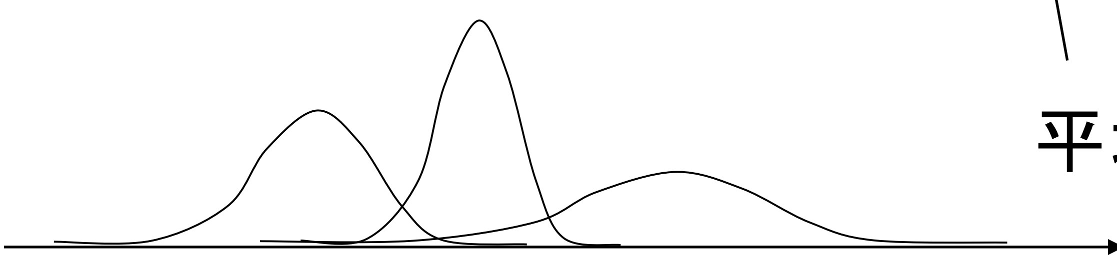
繰り返し法でさがす。

例：混合正規分布

$$w = (a_k, b_k, \sigma_k) \qquad \left(\sum_{k=1}^{K} a_k = 1 \right)$$

$$p(x|w) = \sum_{k=1}^{K} a_k \frac{1}{(2\pi\sigma_k^{\,2})^{N/2}} \exp\left(- \frac{\| x - b_k \|^2}{2\sigma_k^{\,2}} \right)$$

平均 b_k ,分散$\sigma_k^{\,2}$ の
正規分布



隠れ変数(潜在変数)の導入

$$p(x|w) = \sum_{k=1}^{K} a_k \frac{1}{(2\pi\sigma_k^2)^{N/2}} \exp\left(- \frac{\| x - b_k \|^2}{2\sigma_k^2} \right)$$

y について周辺化

$$p(x,y|w) = \prod_{k=1}^{K} \left[a_k \frac{1}{(2\pi\sigma_k^2)^{N/2}} \exp\left(- \frac{\| x - b_k \|^2}{2\sigma_k^2} \right) \right]^{y_k}$$

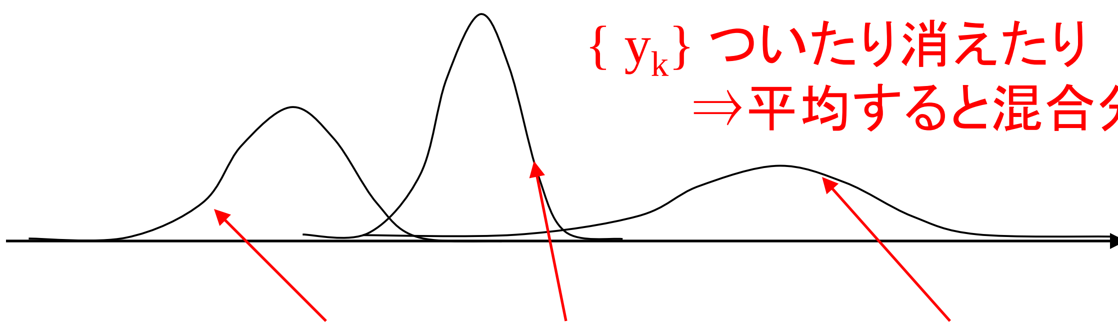
$y = (y_1, y_2, .., y_K)$ は、どれかひとつだけ1で残りは0

隠れ変数を使うと確率分布が簡単になる。

$$p(x,y|w) = \prod_{k=1}^{K} \left[a_k \frac{1}{(2\pi\sigma_k^2)^{N/2}} \exp\left(- \frac{\| x - b_k \|^2}{2\sigma_k^2} \right) \right]^{y_k}$$

$$= \exp\left[- \sum_{k=1}^{K} y_k \{ \|x-b_k\|^2 / 2\sigma_k^2 - N \log \sigma_k + \log a_k \} \right]$$

(定数項は省略)



$\{ y_k \}$ ついたり消えたり
⇒平均すると混合分布

$(y_1,y_2,..,y_K) = (1,0,..,0), (0,1,0,..,0), ..,(0,0,..,1)$

準備

任意の w_1, w_2 について

$$\int p(y|w_1) \log p(y|w_2)\, dy \leqq \int p(y|w_1) \log p(y|w_1)\, dy$$

(証明） カルバック情報量の性質からすぐ。

$$\int p(y|w_1) \log [p(y|w_1)/ p(y|w_2)]\, dx \geqq 0$$

（応用） $\int p(y|w_1) \log p(y|w_2)\, dy$ の w_2 に w_1 を代入すると大きくなる。

EM法のアルゴリズム

$$G(w_1,w_2) = \sum_{i=1}^{n} \sum_y p(y|x_i, w_1) \log p(x_i,y \mid w_2)$$

(1) w_1 初期化
(2) $G(w_1,w_2)$を w_2 について最大化(w_1 固定）
(3) $w_1 := w_2$ として(2)に戻る。

以下で、なぜこれで尤度が単調非減少かを説明する。

なぜEM法でよいのかの説明：$G^*(w_1,w_2)$が必要

定義

$$G^*(w_1,w_2) \equiv G(w_1,w_2) - \sum_{i=1}^{n} \sum_y p(y|x_i,w_1) \log p(y|x_i,w_1)$$

$$= \sum_{i=1}^{n} \sum_y p(y|x_i,w_1) \{ \log p(x_i,y|w_2) - \log p(y|x_i,w_1) \}$$

$$= \sum_{i=1}^{n} \sum_y p(y|x_i,w_1) \{ \log p(x_i|w_2) + \log p(y|x_i,w_2) - \log p(y|x_i,w_1) \}$$

$$= \sum_{i=1}^{n} \log p(x_i|w_2) - \sum_{i=1}^{n} \sum_y p(y|x_i,w_1) \log \frac{p(y|x_i,w_1)}{p(y|x_i,w_2)}$$

関数 $G^*(w_1,w_2)$ は 対数尤度 ー KL情報量

$G^*(w_1,w_2)$

$$= \underbrace{\underbrace{\sum_{i=1}^{n} \log p(x_i \mid w_2)}_{L(w_2)} - \underbrace{\sum_{i=1}^{n} \sum_y p(y \mid x_i, w_1) \log \frac{p(y \mid x_i, w_1)}{p(y \mid x_i, w_2)}}_{w_2=w_1 \text{ のとき最大値0}}}_{}$$

$L(w)$ がw^*で最大$\Leftrightarrow$ $G^*(w_1,w_2)$が$w_1=w_2=w^*$で最大

$L(w)$ を最大化したい$\rightarrow$ $G^*(w_1,w_2)$を最大化すればよい

EM法でなぜ尤度は大きくなるかの説明

$$G^*(w_1,w_2)= G(w_1,w_2) - \sum_{i=1}^{n}\sum_y p(y|\, x_i, w_1) \log p(y|x_i, w_1)$$

$$=\sum_{i=1}^{n}\log p(x_i \mid w_2) - \sum_{i=1}^{n}\sum_y p(y|\, x_i, w_1) \log\frac{p(y|x_i, w_1)}{p(y|x_i, w_2)}$$

(1) w_1 初期化

(2) $G(w_1,w_2)$を w_2 について最大化　$G^*(w_1,w_2)$増加

(3) w_1 ：＝w_2 として(2)に戻る。　$G^*(w_1,w_2)$増加

$G^*(w_1,w_2)$は増加➔$L(w_1)$ が大きくなっていく。

ステップ(2)の最大化の計算を次ページ以下で説明。

ステップ２で最大化する量

$$G(w_1,w_2)=\sum_{i=1}^{n} \sum_y p(y|\, x_i,\, w_1) \log p(x_i,y \mid w_2)$$

をw_2について最大化する。ここで

$$\log p(x_i,{\color{red}y}|w) = -\sum_{k=1}^{K} {\color{red}y_k} \{\|x_i-b_k\|^2/2\sigma_k^2 - N\log\sigma_k + \log a_k\}$$

これの $p(y|\, x_i,\, w_1)$ による平均が必要。

$p(y|\ x_i, w_1)$ による平均は計算できる。

$$\Sigma_y \ \ y_k\, p(y|\ x_i, w)$$

$$= \Sigma_y \ \ y_k\, p(x_i, y|\ w)/p(x_i|w)$$

$$= \frac{a_k \ \ \dfrac{1}{(2\pi\sigma_k{}^2)^{N/2}} \exp\left(- \dfrac{\|\ x_i - b_k \|^2}{2\sigma_k{}^2} \right)}{p(x_i|w_1)}$$

$= E[y_k|\ x_i\ , w_1]$ とかく

ステップ2で最大化する関数を書くと

$$G(w_1,w_2) = \sum_{i=1}^{n} \Sigma_y\, p(y|\, x_i, w_1) \log p(x_i,y \mid w_2)$$

$$= -\sum_{i=1}^{n} \sum_{k=1}^{K} E[y_k|\, x_i\, , w_1]\{\| x_i -b_k\|^2/2\sigma_k{}^2 - N\log\sigma_k + \log a_k\}$$

w_1 が与えられたもとで

これを$w_2=(a_k\, b_k\sigma_k)$について最大化。解ける。

結局、EM法のステップ2は

Σはすべて$i=1,..,n$ の和を表す

$$a_k = \frac{\Sigma E[y_k | x_i, w_1]}{\Sigma 1}$$

$$b_k = \frac{\Sigma E[y_k | x_i, w_1]\, x_i}{\Sigma E[y_k | x_i, w_1]}$$

$$\sigma_k^{\,2} = \frac{\Sigma E[y_k | x_i, w_1]\, \| x_i - b_k \|^2}{N \;\; \Sigma E[y_k | x_i, w_1]}$$

$w_1 \rightarrow w_2$ が計算できる

EM法

(1) w_1 初期化

(2) $w_2=(a_k\ b_k \sigma_k)$を次式で計算

$$
\begin{cases}
a_k = \dfrac{\Sigma\, E[y_k|\, x_i\, , w_1]}{n} \\[2ex]
b_k = \dfrac{\Sigma\, E[y_k|\, x_i\, , w_1]\, x_i}{\Sigma\, E[y_k|\, x_i\, , w_1]} \\[2ex]
\sigma_k^{\ 2} = \dfrac{\Sigma\, E[y_k|\, x_i\, , w_1]\, \|\, x_i - b_k \,\|^{\ 2}}{N\ \ \Sigma\, E[y_k|\, x_i\, , w_1]}
\end{cases}
$$

(3) w_1 ：＝w_2 として(2)に戻る。

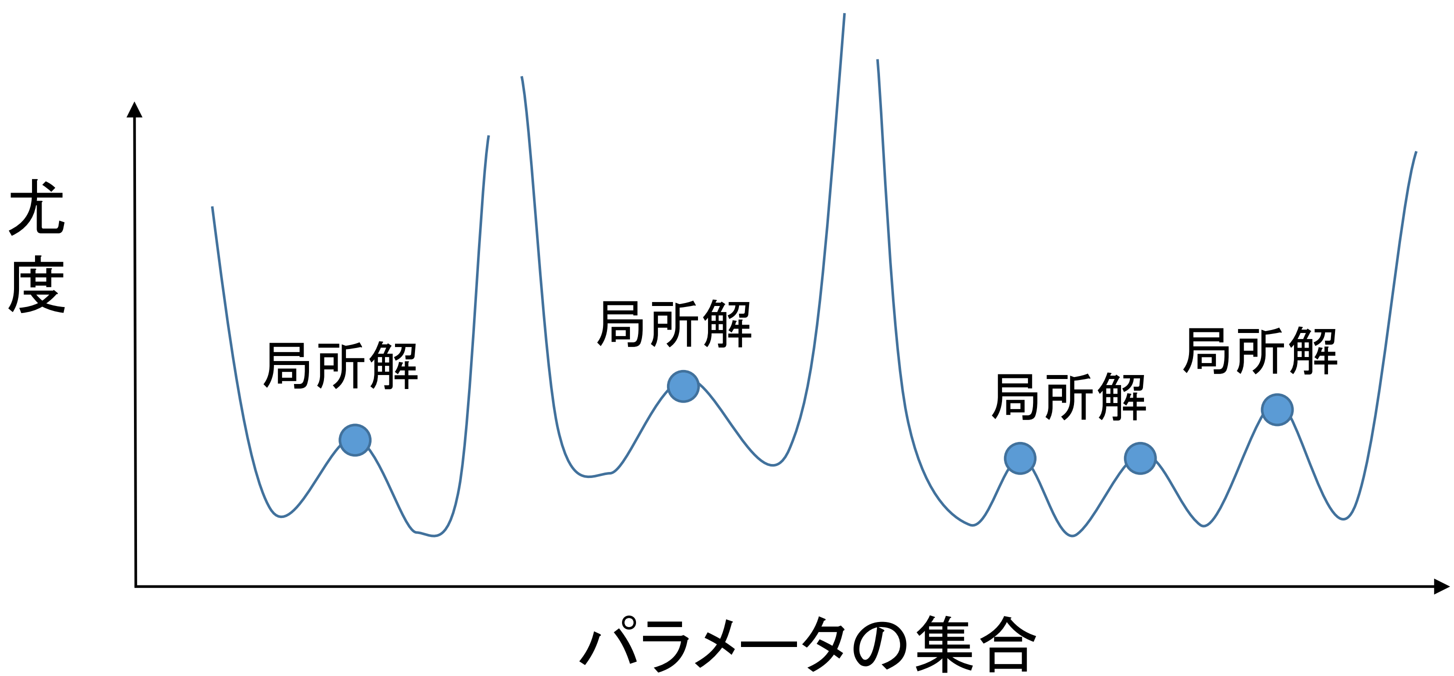
混合正規分布の対数尤度関数の形状

良い局所解を判定することができるだろうか。

(1) 実は混合正規分布では最尤推定量は発散している。($x_i=b_k$, $\sigma_k \to 0$)。

(2) 局所最適解はとてもたくさんあると思われる。

(3) 尤度が大きい局所解が汎化の意味で良いのではないので、
尤度で局所解の良さの比較はできない。AIC, TIC, BIC は使えない。
クロスバリデーション(k-fold)も使えない。
→ 最尤法やEM法は「ダマシダマシ使う」必要がある。
安心して使いたい場合は VB に変更する。EM を VB に変更して
損失になることは何もない(演算量も)。ただし数式は少し難しくなる。



実験例

データ:　$[0,2]^2$ 上の一様分布から発生(n=360)。

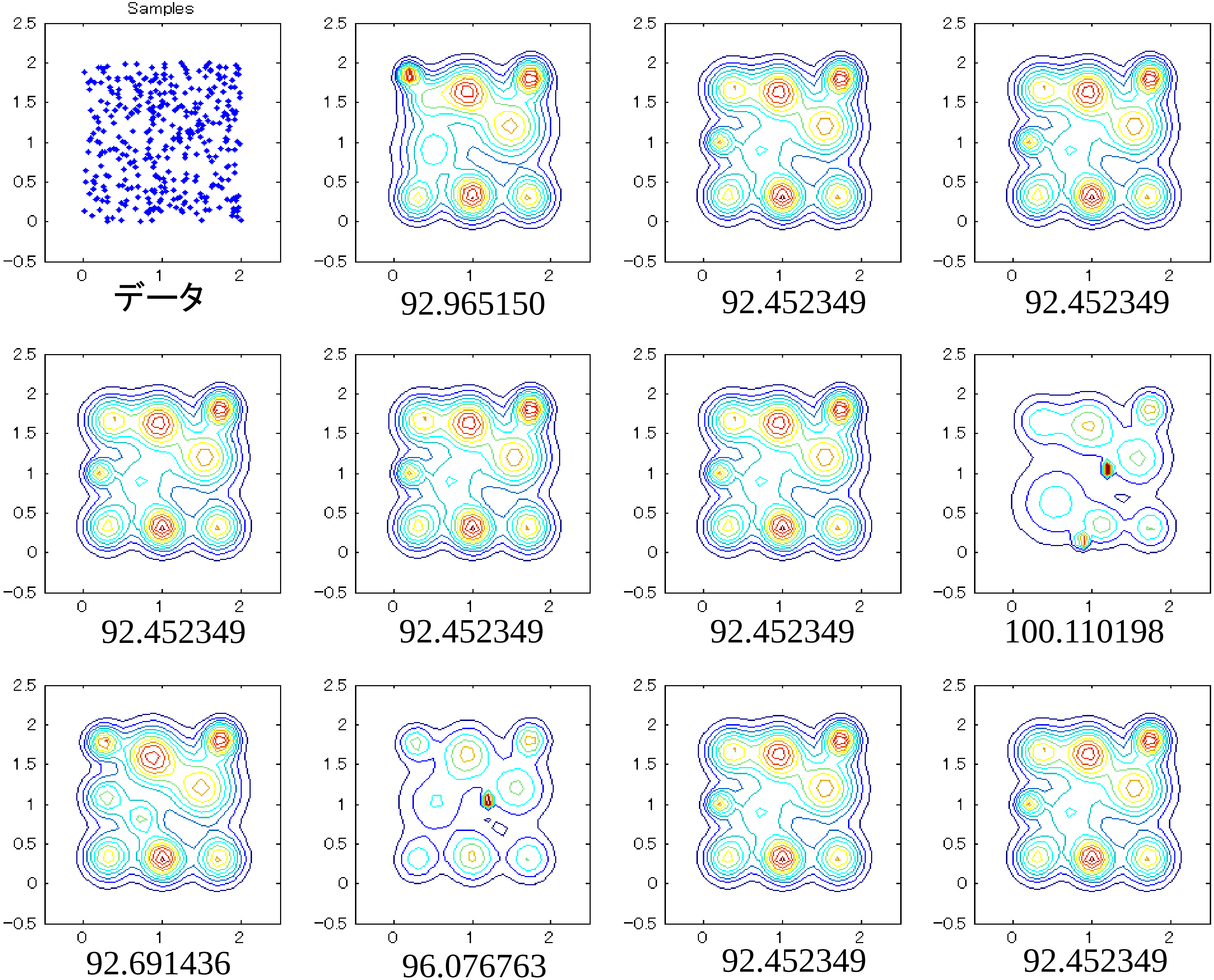
混合正規分布（K=9）で学習（EM法）。

初期値

a_k=1/K,
b_k= 全体平均＋乱数
σ_k= 全体標準偏差

から初めて1000回繰り返し

初期値を変えて11回学習した。尤度が大きい
解は、 σ_kが小さくなっている。


Samples
データ
92.965150
92.452349
92.452349
92.452349
92.452349
92.452349
100.110198
92.691436
96.076763
92.452349
92.452349