統計的機械学習
第5回　混合ガウスモデル1

2013/06/25

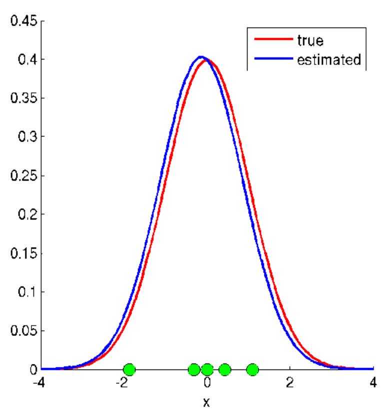
表現工学科　尾形研究室　ゼミ
野田　邦昭

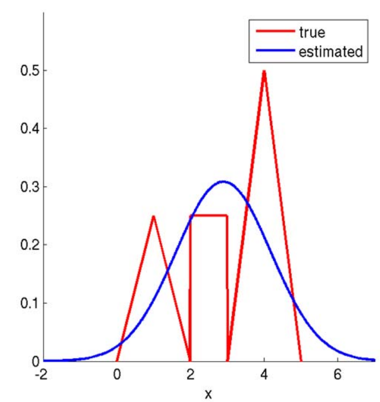
モデルの表現力

ガウスモデル＋最尤推定：

- モデルが（大体）正しい場合，訓練標本数が比較的少なくても推定精度が良い
- モデルが単純なため，表現できる確率密度関数の形が限られる

$$q(x;\mu,\Sigma)=\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$



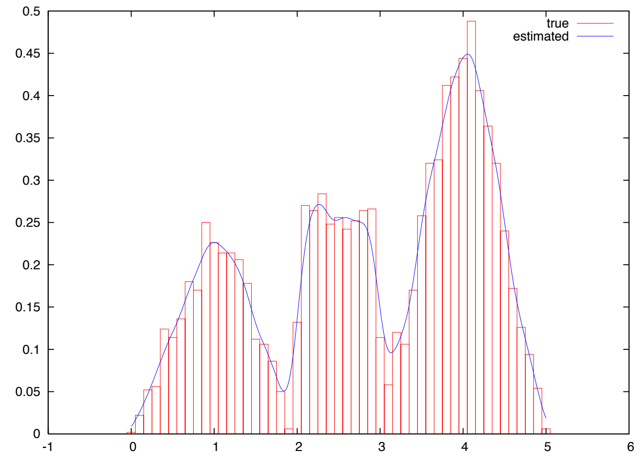


モデルの表現力(続き)

- ガウスカーネル密度推定:
  - 任意の確率密度関数を近似できる
  - 精度良く近似するためには多数の訓練標本が必要

$$\hat{p}(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h} \right)$$

$$K(x) = \frac{1}{(2\pi)^{d/2}} \exp(-\tfrac{1}{2} x^T x)$$



ガウス混合モデル

- ガウス混合モデル(Gaussian mixture model):

$$q(x;\theta)=\sum_{j=1}^{m} w_j \phi(x;\mu_j,\sigma_j)$$

$$w_j \geq 0, \sum_{j=1}^{m} w_j = 1$$

$$\theta=(w_1,\ldots,w_m,\mu_1^T,\ldots,\mu_m^T,\sigma_1,\ldots,\sigma_m)^T$$

$$\mu_j \in \Re^d, \sigma_j > 0$$

$$\phi(x;\mu,\sigma)=\frac{1}{(2\pi\sigma^2)^{d/2}}\exp\left(-\frac{(x-\mu)^T(x-\mu)}{2\sigma^2}\right)$$

m:混合数

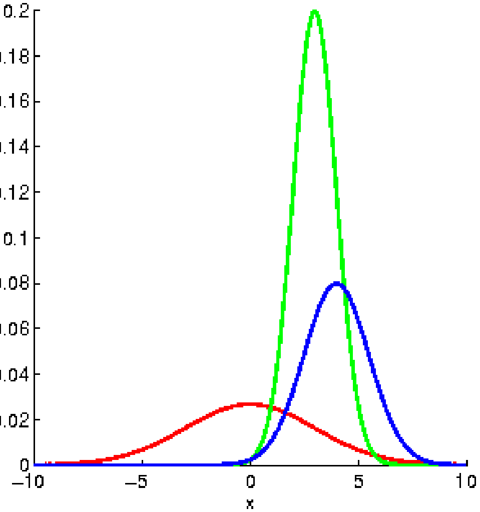
- 注意: $q(x;\theta)$ は確率密度関数なので

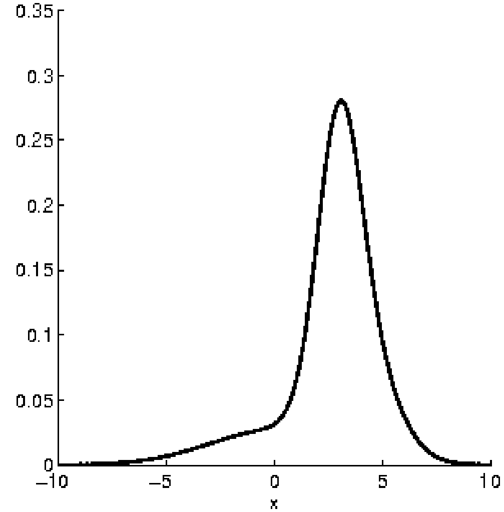
$$\int_D q(x;\theta)dx = 1 \quad \forall x \in D, q(x;\theta) \geq 0$$

を満たす.

ガウス混合モデル

- 有限個のガウスモデルの線形結合:





- 通常のガウスモデルよりも複雑な確率密度関数を表現できる.
- ガウスカーネル密度推定より単純なので，訓練標本が比較的少ない場合でも推定精度が良い.

ガウス混合モデルの最尤推定

- **最尤推定**:(対数)尤度(訓練標本 $\{x_i\}_{i=1}^n$ が生成される確率)を最大にするように θ を定める

$$\log L(\theta) = \sum_{i=1}^n \log q(x_i;\theta) = \sum_{i=1}^n \log \sum_{j=1}^m w_j \phi(x_i;\mu_j,\sigma_j)$$

- 但し，**拘束条件** $w_j \geq 0, \sum_{j=1}^m w_j = 1$ を考慮しなければならない！

- $w_j = \dfrac{\exp(\gamma_j)}{\sum_{j'=1}^m \exp(\gamma_{j'})}$ とおき，$\gamma_j \in \Re$ を決定することにすれば，拘束条件は自動的に満たされる．

尤度方程式

- 最尤推定解の必要条件:

$$\frac{\partial \log L}{\partial \gamma_j} = 0, \quad \frac{\partial \log L}{\partial \mu_j} = 0, \quad \frac{\partial \log L}{\partial \sigma_j} = 0$$

より，最尤推定解は以下を満たす（証明は宿題）

$$\hat{w}_j = \frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_{i,j}$$

$$\hat{\sigma}_j = \sqrt{\frac{1}{d}\frac{\sum_{i=1}^{n}\hat{\eta}_{i,j}(x_i - \hat{\mu}_j)^T(x_i - \hat{\mu}_j)}{\sum_{i'=1}^{n}\hat{\eta}_{i',j}}}$$

$$\hat{\mu}_j = \frac{\sum_{i=1}^{n}\hat{\eta}_{i,j}x_i}{\sum_{i'=1}^{n}\hat{\eta}_{i',j}}$$

$$\hat{\eta}_{i,j} = \frac{\hat{w}_j\phi(x_i;\hat{\mu}_j,\hat{\sigma}_j)}{\sum_{j'=1}^{m}\hat{w}_{j'}\phi(x_i;\hat{\mu}_{j'},\hat{\sigma}_{j'})}$$

$d : x$の次元

- しかし，この連立方程式は簡単に解けない．

勾配法

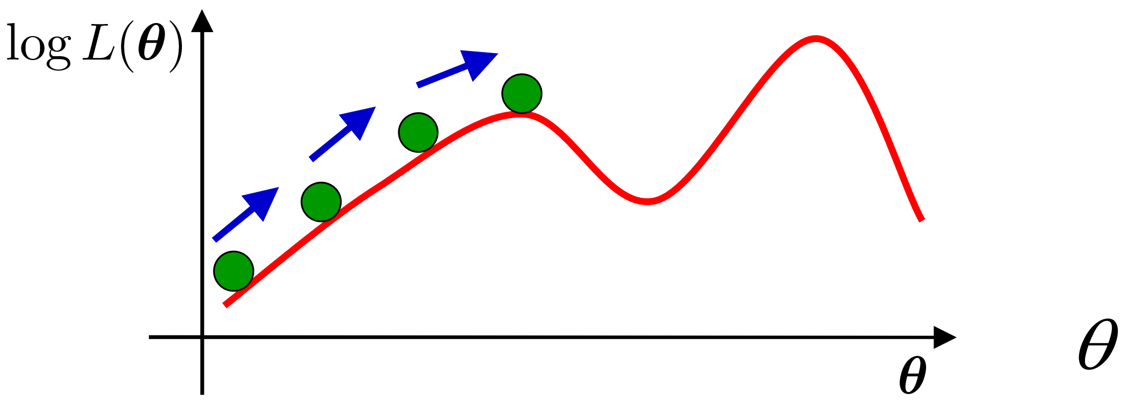
- 勾配法(gradient method):
  1. 適当に初期値 $\hat{\theta}^{(0)}$ を定める.
  2. 勾配を上るようにパラメータを更新する:

$$\hat{\theta}^{(t+1)} \leftarrow \hat{\theta}^{(t)} + \varepsilon \left.\frac{\partial \log L}{\partial \theta}\right|_{\theta=\hat{\theta}^{(t)}}$$

ε :小さい正のスカラー

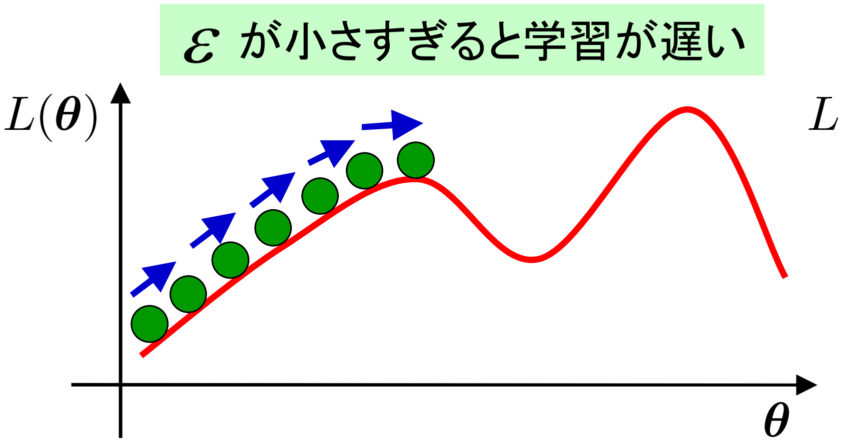
  3. 収束するまで勾配上昇を繰り返す.

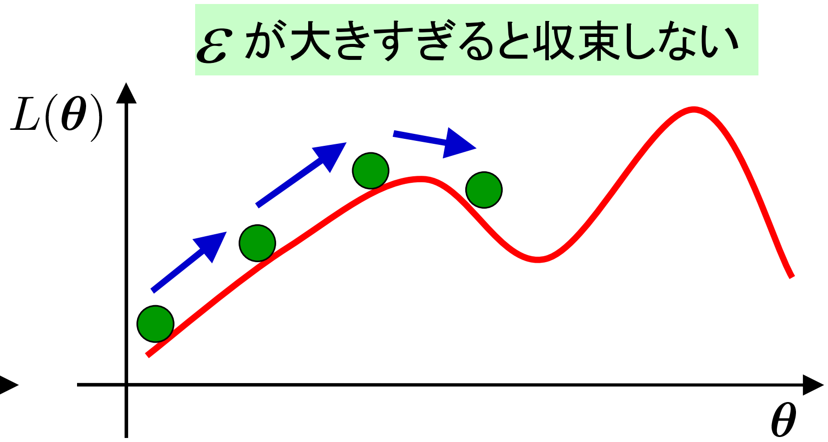
- 局所最適解(local optimal solution)が求まる.



勾配法の欠点

■ 学習率 ε の選び方が難しい：





- ε は，「最初は大きく，徐々に小さく」が原則だが，適切に決定することは難しい

■ 局所最適解しか見つけられない：

- 様々な初期値から何度か学習し，最適な値を採用する