２．３．９　混合ガウス分布

7月4日

田中洸一

手順

- 問題提起
- 混合ガウス分布
- 負担率
- パラメータの解法

問題提起（1）

基本的な分布を重ね合わせ、

混合分布という確率モデルを作る

↓

重ね合わせの重み係数、各分布の平均、

共分散を調整することで、任意の連続な密度関数を表現できる

問題提起(2)

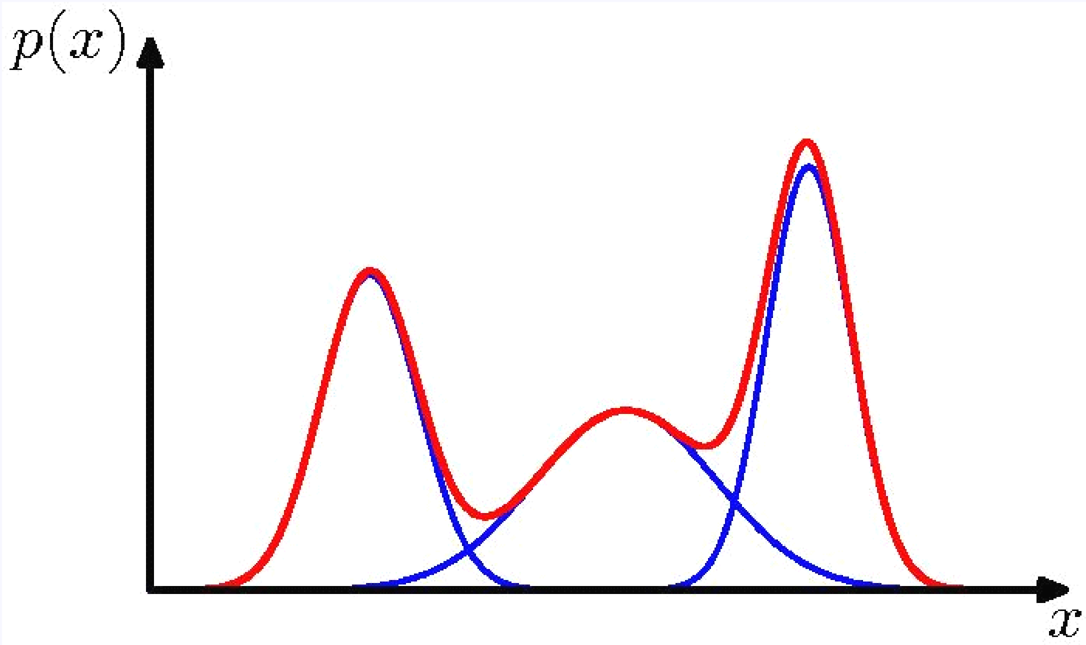
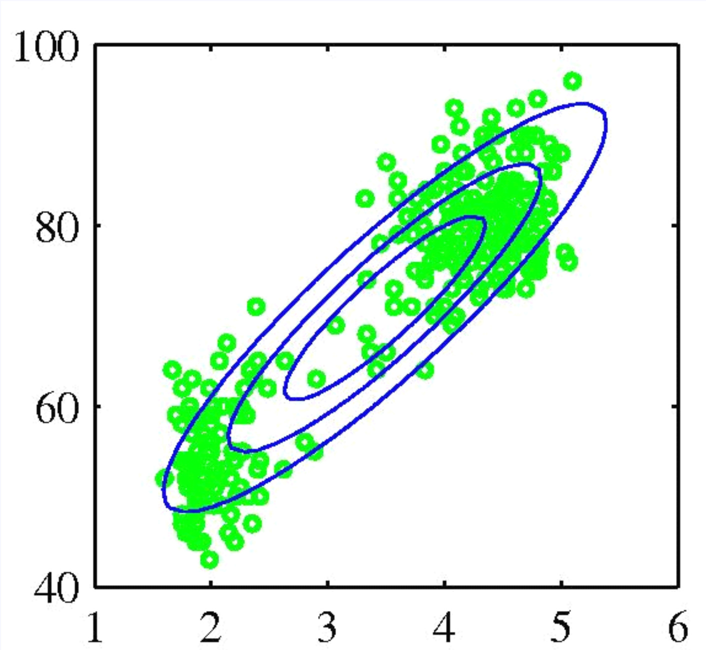


図1:3つのガウス分布(青線)、
それらの和　(赤)

問題提起（3）

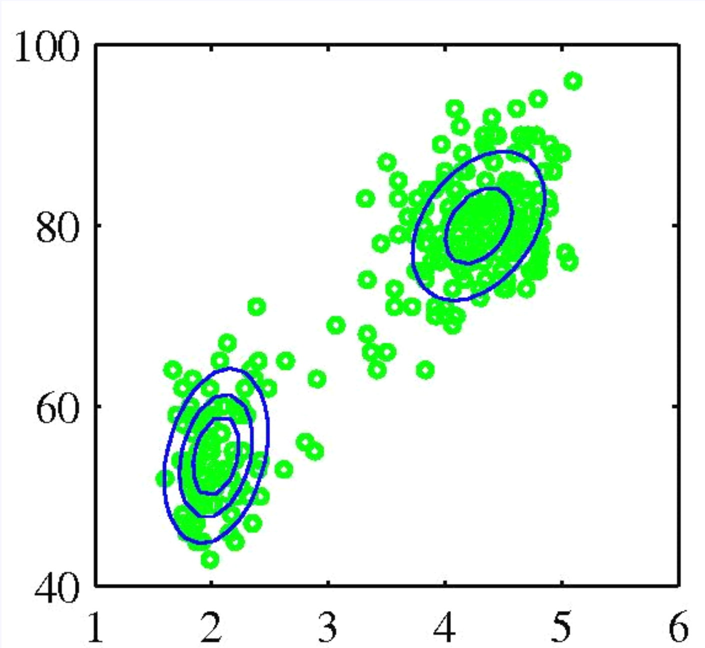


単一のガウス分布から最尤推定で

フィッティング

→ うまく捉えてない

問題提起(3)



最尤推定でフィッティングした、

２つのガウス分布の線形結合である分布

→　うまく捉えられている

混合ガウス分布（1）

- K個のガウス分布の重ね合わせ

$$p(\mathrm{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathrm{x} \mid \mu_k, \Sigma_k) \qquad (2.188)$$

混合係数：π_k

混合要素：$\mathcal{N}(\mathrm{x} \mid \mu_k, \Sigma_k)$

※各ガウス分布個別の平均μ_kと共分散Σ_kを持つ

混合係数

パラメータ π_k について、p(x)と各ガウス分布が正規化されているので、(2.188) の両端をxについて積分

$$\sum_{k=1}^{k} \pi_k = 1 \qquad (2.189)$$

p(x) ≥ 0 より、すべてのkについて $\pi_k \geq 0$ つまり、

$$0 \leq \pi_k \leq 1 \qquad (2.190)$$

混合係数を確率としてみることができる

混合ガウス分布（２）

π_k :k番目の混合要素を選択する事前分布 $\mathrm{p}(k)$

$\mathcal{N}(\mathrm{x} \mid \mu_k, \Sigma_k)$:kが与えられた時のxの条件付き密度 $\mathrm{p}(\mathrm{x} \mid k)$

確率の加法・乗法定理から、xの周辺密度

$$\mathrm{p}(\mathrm{x}) = \sum_{k=1}^{K} \mathrm{p}(k)\,\mathrm{p}(\mathrm{x} \mid k) \qquad (2.191)$$

$$\mathrm{p}(\mathrm{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathrm{x} \mid \mu_k, \Sigma_k) \qquad (2.188)$$

２つの式が等しい

負担率

- 事後確率 $p(k|x)$ は負担率として知られ、重要な役割を果たす。

$$
\begin{aligned}
\Upsilon_k(x) &\equiv p(x|k) \\
&= \frac{p(k)p(x|k)}{\sum_l p(l)p(x|l)} \\
&= \frac{\pi_k \mathcal{N}(x|\mu_k,\Sigma_k)}{\sum_l \pi_l \mathcal{N}(x|\mu_l,\Sigma_l)} \qquad (2.192)
\end{aligned}
$$

パラメータの解法（１）

混合ガウス分布を表現するには、

以下の<u>パラメータ</u>によって決まる

$$\pi \equiv \{\pi_1, \ldots, \pi_K\}$$

$$\mu \equiv \{\mu_1, \ldots, \mu_K\}$$

$$\Sigma \equiv \{\Sigma_1, \ldots, \Sigma_K\}$$

パラメータの値を求める　→　最尤推定法

パラメータの解法（2）

パラメータの値を求める　→　最尤推定法

（2．188）より、

$$X = \left\{ x_1, \ldots, x_K \right\}$$ についての対数尤度関数

$$p(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x \mid \mu_k, \Sigma_k) \qquad (2.188)$$

$$\ln p(X \mid \pi, \mu, \Sigma) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(x_n \mid \mu_k, \Sigma_k) \right\} \qquad (2.193)$$

尤度関数が<u>最大となるパラメータ</u>を求める

パラメータの解法（3）

(2.193) が最大となるパラメータを求める・・・が、
式が複雑で普通には解は得られない

↓

尤度関数を最大化するアプローチ法

・繰り返し的な数値最適化手法

・<u>EMアルゴリズム</u>