混合分布のパラメータ推定

本：わかりやすいパターン認識

菊池裕紀

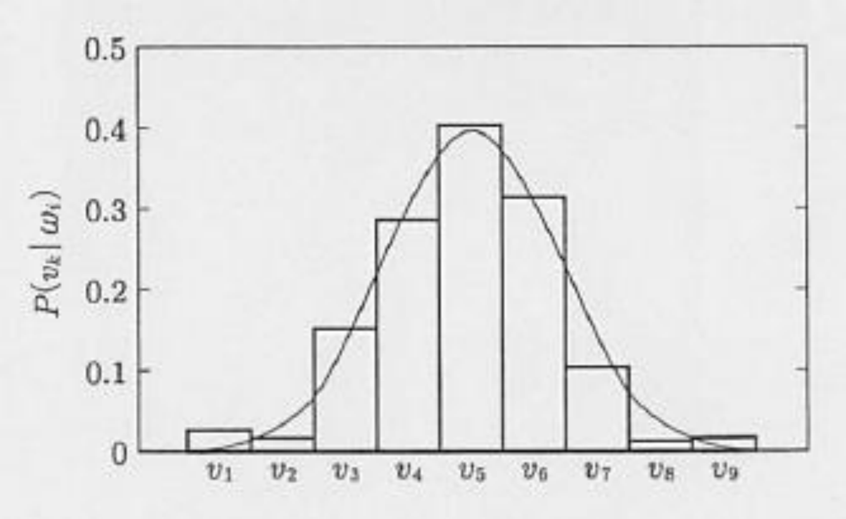
混合分布に対するパラメトリックな手法

確率密度関数

$p(\boldsymbol{x}|\omega_i) \qquad (\boldsymbol{x} = \{v_1, v_2, \ldots, v_m\})$

↓ **正規分布に従うとすると…**

$$p(x|\omega_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{\frac{1}{2}(x-\mu_i)^2}{\sigma_i^2}\right]$$



× $P(v_k|\omega_i) \quad (k = 1,2,\ldots,m)$**を推定する**

○x **の平均**μ_i**と分散**σ_i^2**の2つのパラメータを求める**（**パラメトリックな学習**）

観測データから直接識別関数を設計（ノンパラメトリックな学習）

パラメトリック手法によるパラメータ推定

- クラスω_i $(i = 1,2,\dots,c)$ がそれぞれ異なった確率密度比をもつ
- 観測結果が c 個の確率密度比による**混合分布**に従う
- 観測データは独立でマルコフ性はない

$\rightarrow$ パラメトリックな手法によるパラメータ推定

ω_i の確率密度関数 $p(\boldsymbol{x}|\omega_i;\theta_i)$

$\theta = (\theta_1,\theta_2,\dots,\theta_c,\ \pi_1,\pi_2,\dots,\pi_c)$

混合分布の確率密度関数 $p(x_k;\theta) = \sum_{i=1}^{c} \pi_i * p(x_k|\omega_i;\boldsymbol{\theta}_i) \quad (k = 1,2,\dots,n)$

パラメータ推定を「教師付き学習」「教師なし学習」で行っていく

教師付き学習によるパラメータ推定（１）

- **教師付き学習の場合：**

データ： $(x_1, s_1), (x_2, s_2), \dots, (x_n, s_n) \quad s_k \in (w_1, w_2, \dots, w_c)$

観測結果と所属クラスが対の形のデータが得られる

尤度： $p(\boldsymbol{X}, \boldsymbol{S}; \boldsymbol{\theta}) = \prod_{\boldsymbol{k=1}}^{\boldsymbol{n}} P(s_k) \prod_{\boldsymbol{k=1}}^{\boldsymbol{n}} p(x_k | s_k; \theta_1, \theta_2, \dots, \theta_c)$

（$\boldsymbol{X} = (x_1, x_2, \dots, x_n)$，$\boldsymbol{S} = (s_1, s_2, \dots, s_n)$）

対数尤度： $\log p(\boldsymbol{X}, \boldsymbol{S}, \boldsymbol{\theta}) = \sum_{k=1}^{n} \log P(s_k) + \sum_{k=1}^{n} \log p(x_k | s_k; \theta_1, \dots, \theta_c)$

$= L_1 + L_2$

教師付き学習によるパラメータ推定(２)

$\log p(\mathbf{X}, \mathbf{S}, \boldsymbol{\theta}) = \mathrm{L}_1 + \mathrm{L}_2$

$L_1 \overset{\text{def}}{=} \sum_{k=1}^{n} \log P(s_k)$：

π_iについて最大化

$L_2 \overset{\text{def}}{=} \sum_{k=1}^{n} \log p(x_k | s_k; \theta_1, \dots, \theta_c)$：

θ_iについて最大化

L_1の最大化：

$\widehat{\pi}_i = \frac{n_i}{n}$ (n_i:ω_iから発生した観測データの総数)

教師付き学習によるパラメータ推定（２）

- L_2の最大化：

$$L_2 = \sum_{i=1}^{c} \sum_{x_k \in \omega_i} \log\ p(x_k | w_i; \theta_i)$$

これを最大化するθ_iは以下を満たす必要がある：

$$\nabla_{\theta_i} L_2 = \sum_{x_k \in \omega_i} \nabla_{\theta_i} \log p(x_k | w_i; \theta_i) = 0$$

教師なし学習によるパラメータ推定（１）

- 教師なし学習の場合：

  データ：　$(x_1, x_2, \dots, x_n)$

  対数尤度：　$\log p(\boldsymbol{X}; \boldsymbol{\theta}) = \sum_{k=1}^{n} \log p(x_k; \boldsymbol{\theta})$

  制約条件：　$\sum_{i=1}^{c} \pi_i = 1$

$$L = \log p(X; \theta) - \lambda\left(\sum_{i=1}^{c} \pi_i - 1\right)$$

$$\frac{\partial L}{\partial \pi_i} = 0, \qquad \nabla_{\theta_i} L = 0$$

教師なし学習によるパラメータ推定（２）

- π_iの推定

$$\widehat{\pi_i} = \frac{1}{n}\sum_{k=1}^{n} P(\omega_i | x_k; \boldsymbol{\theta})$$

- θ_iの推定

$$\begin{aligned}
\nabla_{\theta_i}\mathrm{L} &= \nabla_{\theta_i}\log p(\boldsymbol{X}; \boldsymbol{\theta}) \\
&= \textstyle\sum_{k=1}^{n} \frac{1}{p(x_k;\theta)} \nabla_{\theta_i} p(x_k; \boldsymbol{\theta}) \\
&= \textstyle\sum_{k=1}^{n} \frac{1}{p(x_k;\theta)} \nabla_{\theta_i}\left(\sum_{j=1}^{c} \pi_j * p(x_k | w_j; \theta_j)\right) \\
&= \textstyle\sum_{k=1}^{n} \frac{\pi_i}{p(x_k;\theta)} \nabla_{\theta_i} p(x_k | \omega_i; \theta_i) = 0
\end{aligned}$$

教師なし学習によるパラメータ推定（３）

ベイズ推定を利用：

$$P(\omega_i|x_k;\theta) = \frac{\pi_i * p(x_k|\omega_i;\theta_i)}{p(x_k;\theta)}$$

$$\nabla_{\theta_i} L = \sum_{k=1}^{n} P(\omega_i|x_k;\theta) * \frac{\nabla_{\theta_i} p(x_k|\omega_i;\theta)}{p(x_k|\omega_i;\theta_i)}$$

$$= \sum_{k=1}^{n} P(\omega_i|x_k;\theta) \nabla_{\theta_i} \log\, p(x_k|\omega_i;\boldsymbol{\theta}_i) = 0$$

推定手順

1. 事前確率π_iおよびパラメータθ_iに初期値を与える
2. ベイズの定理より、各x_kに対して$P(\omega_i|x_k;\theta)$を計算する
3. 以下の式によりπ_i, θ_iを更新し、新しい値を得る
   - $\hat{\pi}_i = \frac{1}{n}\sum_{k=1}^{n} P(\omega_i|x_k;\theta)$
   - $\sum_{k=1}^{n} P(\omega_i|x_k;\theta)\nabla_{\theta_i}\log p(x_k|\omega_i;\theta_i) = 0$
4. 新しい値で対数尤度を計算し、その増分が閾値以下なら終了し、違えば Step2に戻る

混合正規分布のパラメータ推定

- 混合分布の確率密度関数に正規分布を考える
- パラメータを教師なし学習で推定

$$p(\mathrm{X_k}|\omega_i;\theta_i) = \frac{1}{(2\pi)^{d/2}|\Sigma_i|}\exp\left[-\frac{1}{2}(X_k-\mu_i)^t\Sigma_i^{-1}(X_k-\mu_i)\right]$$

$(\theta_i = (\mu_i, \Sigma_i))$

$$\widehat{\mu_i} = \frac{\sum_{k=1}^{n} P(\omega_i|X_k;\theta)X_k}{\sum_{k=1}^{n} P(\omega_i|X_k;\theta)} \quad ,\widehat{\Sigma_i} = \frac{\sum_{k=1}^{n} P(\omega_i|X_k;\theta)(X_k-\widehat{\mu_i})(X_k-\widehat{\mu_i})^t}{\sum_{k=1}^{n} P(\omega_i|X_k;\theta)}$$

$$\widehat{\pi}_i = \frac{1}{n}\sum_{k=1}^{n} P(\omega_i|X_k;\theta)$$

推定手順

1. **事前確率**π_i**および**μ_i, Σ_i**の初期値を与える**
2. **ベイズの定理より、各**X_k**に対して**$P(\omega_i|X_k;\theta)$**を計算**
3. **以下より新しい値を求める**

   □$\widehat{\pi}_i = \frac{1}{n}\sum_{k=1}^{n} P(\omega_i|X_k;\theta)$，$\widehat{\mu}_i = \frac{\sum_{k=1}^{n} P(\omega_i|X_k;\theta)X_k}{\sum_{k=1}^{n} P(\omega_i|X_k;\theta)}$

   □$\widehat{\Sigma}_i = \frac{\sum_{k=1}^{n} P(\omega_i|X_k;\theta)(X_k-\widehat{\mu_i})(X_k-\widehat{\mu_i})^t}{\sum_{k=1}^{n} P(\omega_i|X_k;\theta)}$

4. **対数尤度を求め、増分が設定した閾値以下なら終了し、違うならば**Step2**へ**