ディリクレ過程混合モデルによるクラスタリング

続・わかりやすいパターン認識第 12 章

Satoshi Murashige

October 8, 2018

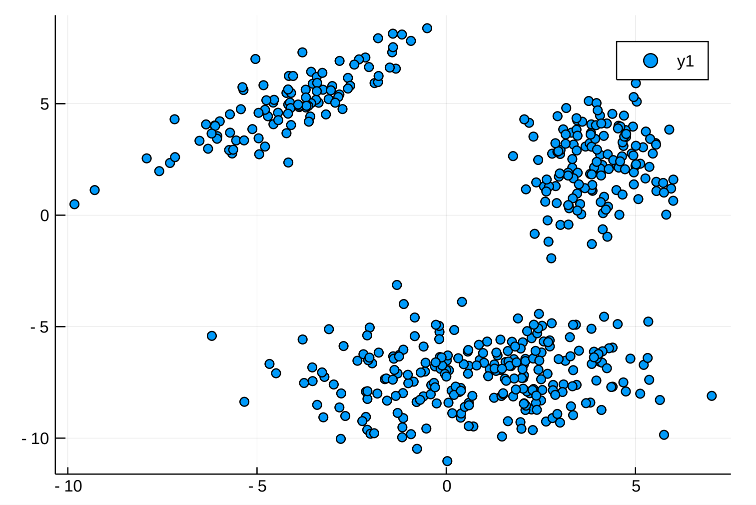
Mathematical Informatics Lab., NAIST

Title

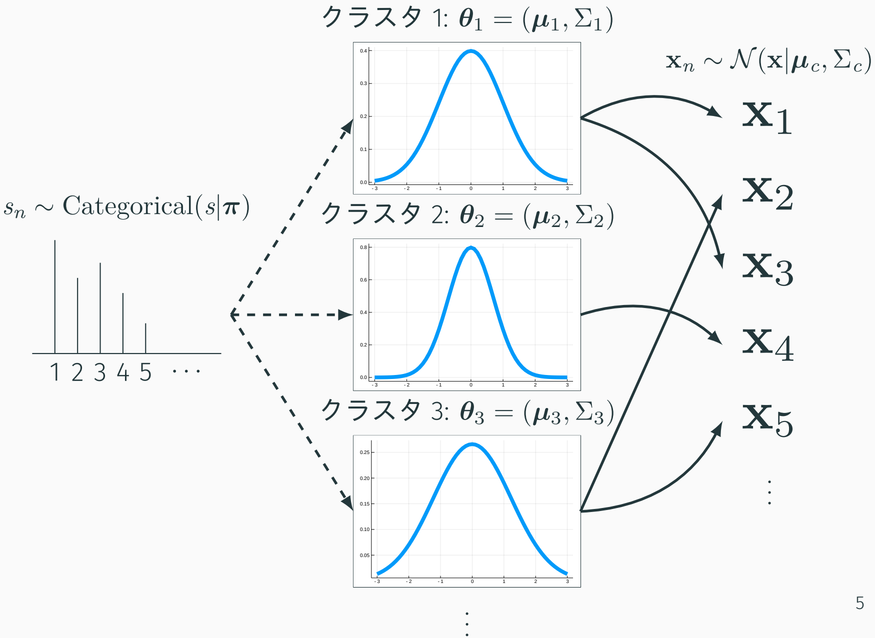
Hello metropolis!

Table of Contents

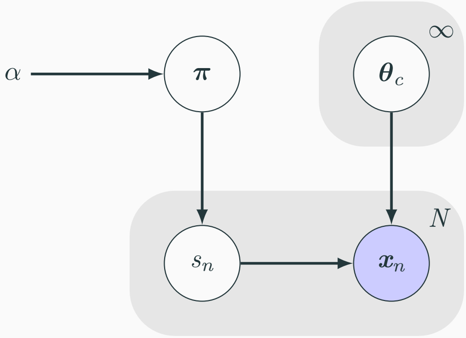
モチベーション：
クラスタ数が未知のクラスタリング問題を解きたい



生成モデル：観測パターンの生成過程をモデル化する



観測パターンの生成過程のグラフィカルモデル



クラスタリング問題の定式化

$$p(\mathbf{s}, \boldsymbol{\theta}|\mathbf{x}) = \frac{p(\mathbf{x}, \mathbf{s}, \boldsymbol{\theta})}{p(\mathbf{x})} = \frac{p(\mathbf{x}|\mathbf{s}, \boldsymbol{\theta})P(\mathbf{s})p(\boldsymbol{\theta})}{\sum_{\mathbf{s}} \int p(\mathbf{x}|\mathbf{s}, \boldsymbol{\theta})P(\mathbf{s})p(\boldsymbol{\theta})d\boldsymbol{\theta}} \tag{11.4}$$

$$(\hat{\mathbf{s}}, \hat{\boldsymbol{\theta}}) = \underset{\mathbf{s}, \boldsymbol{\theta}}{\operatorname{argmax}} \{p(\mathbf{s}, \boldsymbol{\theta}|\mathbf{x})\} \tag{11.5}$$

1. 事前確率 $P(\mathbf{s})$ をどのように決めるか？
    - クラスタ数が無限にあるので，工夫が必要
2. 事後確率 $p(\mathbf{s}, \boldsymbol{\theta}|\mathbf{x})$ をどのように最大化するか？
    - $\mathbf{s}$ のあらゆる組合せを考えるのは不可能

事前確率 $P(\mathrm{s})$ はデータをクラスタに分割する方法の確率モデル

クラスタリング	CRP
k 個目のパターン $\mathbf{x}_k$	k 人目の客
総パターン数 n	来店客数
クラスタ ω_i	テーブル i
クラスタ数 c	使用するテーブル数
クラスタ ω_i のパターン数 n_i	テーブル i に座った人の数

クラスタ数未知のクラスタリング問題におけるデータの生成過程の定式化

$$s_1, \cdots, s_n | \alpha \sim \mathrm{CRP}(\alpha) \tag{12.4}$$

$$\boldsymbol{\theta}_i | \beta \sim G_0(\boldsymbol{\theta}) \tag{12.5}$$

$$\mathbf{x}_k | s_k = \omega_i, \theta \sim p(\mathbf{x} | \boldsymbol{\theta}_i) \tag{12.6}$$

事後確率 $p(\mathrm{s}, \boldsymbol{\theta} | \mathrm{x})$ の最大化問題を解きたい

$$(\hat{\mathrm{s}}, \hat{\boldsymbol{\theta}}) = \underset{\mathbf{s}, \boldsymbol{\theta}}{\operatorname{argmax}} \{ p(\mathbf{s}, \boldsymbol{\theta} | \mathbf{x}) \} \tag{11.5}$$

アプローチ

- あらゆる組合せの s を評価するのではなく，ランダムサンプリングした s の中から $p(\mathrm{s}, \boldsymbol{\theta} | \mathrm{x})$ を最大化するものを選ぶ
- ランダムサンプリングには Gibbs サンプリングを用いる

事後確率最大化のアルゴリズム

1. aaa
2. aaa
3. aaa
4. aaa
5. aaa

Second Frame

$$G(\boldsymbol{\theta})|\alpha, G_0(\boldsymbol{\theta}) \sim \mathrm{DP}(\alpha, G_0(\boldsymbol{\theta})) \tag{12.1}$$

$$\boldsymbol{\theta}^k \sim G(\boldsymbol{\theta}) \qquad (k = 1, \cdots, n) \tag{12.2}$$

$$\mathbf{x}_k|\boldsymbol{\theta} \sim p(\mathbf{x}|\boldsymbol{\theta}^k) \qquad (k = 1, \cdots, n) \tag{12.3}$$

Summary

Appendix

データの生成過程の具体例：ニュース記事の生成過程

$$G(\boldsymbol{\theta})|\alpha, G_0(\boldsymbol{\theta}) \sim \mathrm{DP}(\alpha, G_0(\boldsymbol{\theta})) \tag{12.1}$$

$$\boldsymbol{\theta}^k \sim G(\boldsymbol{\theta}) \qquad (k = 1, \cdots, n) \tag{12.2}$$

$$\mathbf{x}_k|\boldsymbol{\theta} \sim p(\mathbf{x}|\boldsymbol{\theta}^k) \qquad (k = 1, \cdots, n) \tag{12.3}$$

Dirichlet 過程混合モデルによるデータの生成過程の記述

$$G(\boldsymbol{\theta})|\alpha, G_0(\boldsymbol{\theta}) \sim \mathrm{DP}(\alpha, G_0(\boldsymbol{\theta})) \tag{12.1}$$

$$\boldsymbol{\theta}^k \sim G(\boldsymbol{\theta}) \qquad (k = 1, \cdots, n) \tag{12.2}$$

$$\mathbf{x}_k|\boldsymbol{\theta} \sim p(\mathbf{x}|\boldsymbol{\theta}^k) \qquad (k = 1, \cdots, n) \tag{12.3}$$

Remark: Chinese Restaurant Process（CRP) は客がテーブルを選択する過程をモデル化したもの

$$s_1, \cdots s_n|\alpha \sim \mathrm{CRP}(\alpha) \tag{12.4}$$

$$\boldsymbol{\theta}_i|\beta \sim G_0(\boldsymbol{\theta}) \qquad (i = 1, \cdots, \infty) \tag{12.5}$$

$$\mathbf{x}_k|s_k = \omega_i, \boldsymbol{\theta} \sim p(\mathbf{x}|\boldsymbol{\theta}_i) \qquad (k = 1, \cdots, n) \tag{12.6}$$

- CRP では，無限個のテーブルがある料理店を考えているので，クラスタの数が無限にある.
- α は新しいテーブルが選ばれる確率

Bayes 推定に基づくクラスタリングのモデル化：クラスタ割当てとクラスタのモデルパラメタが未知

- クラスタリング問題の確率モデルを立てる

$$\text{パターン}:\mathbf{x}=\{\mathbf{x}_1,\cdots,\mathbf{x}_n\} \tag{11.1}$$

$$\text{パターンのクラスタ割当て}:\mathbf{s}=\{s_1,\cdots,s_n\} \tag{11.2}$$

$$\text{各クラスタのモデルパラメタ}:\boldsymbol{\theta}=\{\boldsymbol{\theta}_1,\cdots,\boldsymbol{\theta}_c\} \tag{11.3}$$

- Bayes の定理から事後確率を計算する

$$p(\mathbf{s},\boldsymbol{\theta}|\mathbf{x})=\frac{p(\mathbf{x}|\mathbf{s},\boldsymbol{\theta})P(\mathbf{s})p(\boldsymbol{\theta})}{p(\mathbf{x})}=\frac{p(\mathbf{x}|\mathbf{s},\boldsymbol{\theta})P(\mathbf{s})p(\boldsymbol{\theta})}{\sum_{\mathbf{s}}\int p(\mathbf{x}|\mathbf{s},\boldsymbol{\theta})P(\mathbf{s})p(\boldsymbol{\theta})\,d\boldsymbol{\theta}} \tag{11.4}$$

- 事後確率最大化によりクラスタ割当て・パラメタを決定

$$(\hat{\mathbf{s}},\hat{\boldsymbol{\theta}})=\underset{\mathbf{s},\boldsymbol{\theta}}{\operatorname{argmax}}\{P(\mathbf{s},\boldsymbol{\theta}|\mathbf{x})\} \tag{11.5}$$