

目次 前回 次回 略解

理論物理学特論

樋口さぶろお*¹ 配布: 2012-07-13 Fri 更新: Time-stamp: "2012-07-13 Fri 07:28 JST hig"

11 クラスタ分析

11.1 略解: クラスタ分析

1. 距離 2 で, $\{\{A,B\},\{C\},\{D\}\}$. 距離 4 で, $\{\{A,B\},\{C,D\}\}$. 距離 5 で, $\{\{A,B,C,D\}\}$.
2. クラスタ $\{A,B\}$ の座標の平方偏差和は 2. クラスタ $\{C,D\}$ の座標の平方偏差和は 4. クラスタ $\{A,B,C,D\}$ の座標の平方偏差和は 40. よって, 距離は 34.
3. 2 個の (空でない) クラスタへの分割方法は 7 個あるが, その中で最小となるのは $\{A,B\},\{C,D\}$. 目的関数の値は $2 + 4 = 6$.
4. ステップ 0 でのクラスタは $C_1 = \{A,B,C\}, C_2 = \{D\}$. それぞれの重心は $(\frac{8}{3}, \frac{7}{3}), (6, 7)$
ステップ 1 でのクラスタは $C_1 = \{A,B\}, C_2 = \{C,D\}$.

12 クラスタ分析 (2)

今日の目標

- k -means のアルゴリズムが書ける

12.1 quiz: クラスタ分析

2 変量データ $A(1,1), B(3,1), C(4,5), D(6,7)$ を考える.

1. クラスタ間距離を, 最近接データ点間のマンハッタン距離, としたとき階層的クラスタ分析を行い, デンドログラムを描こう.
2. クラスタ (A,B) と (C,D) の間の, 平方偏差和の増分で定義される距離 (ウォードの方法で用いられる) を計算しよう.
3. 目的関数を, クラスタ内データ間のマンハッタン距離の最大, のすべてのクラスタにわたる和, とする. このとき非階層的クラスタ分析を行って 2 個のクラスタに分

*¹ Copyright ©2012 Saburo HIGUCHI. All rights reserved.

hig@math.ryukoku.ac.jp, <http://hig3.net>(講義のページもここからたどれます), へや:1 号館 5 階 502.

けよう.

4. k -means 法で $k = 2$ とする. シードを, $(4, 4), (5.7)$ とするとき, 最初の 2 回の繰り返しでのクラスターを求めよう.

R Commander **メニュー探検**

- 統計量 > 次元解析 > クラスタ分析 > 階層的クラスタ分析
- 統計量 > 次元解析 > クラスタ分析 > k -平均クラスタ分析

R Console で, stats ライブラリの kmeans 関数. kmeans(データ, クラスタ数).

12.2 レポート計画

■**成績の計算** 科目の成績 100 ピーナッツは

- 30 ピーナッツ:いろいろ (=Quiz 14 ピーナッツ + **レポート 16 ピーナッツ**)
- 30 ピーナッツ:プチテスト いまのところ 2012-06-15 を予定
- 40 ピーナッツ:ファイナルトライアル

Web その他からとってきたデータを, この授業で学んだ手法の 1 個以上を使って分析して, Mahara のページにまとめてください. 複数の参加者が同じデータにならないように調整.

理論物理学特論のグループで共有し, 樋口に「送信」してください.

期限 2012-08-03 金 23:55(ファイナルトライアルの日)