非/階層的クラスタリング

樋口さぶろお https://hig3.net

龍谷大学 先端理工学部 数理・情報科学課程

多変量解析☆演習　L09(2026-11-19 Thu)

最終更新: Time-stamp: "2026-11-19 Thu 16:11 JST hig"

今日の目標

- 非/階層的クラスタ分析の違いを説明できる
- 主なクラスタリングアルゴリズムを説明できる

ここまで来たよ

9. 非/階層的クラスタリング
   - 様々なクラスタリングのアルゴリズム
   - （モデルを仮定しない）非階層的クラスタリング：K-means 法
   - 階層的クラスタリング, 距離に基づく, Ward 法
   - クラスタリングの評価
   - 対称行列の固有値・固有ベクトル・対角化

クラスタリングのアルゴリズム

クラスタリング＝「近い」データ点をクラスターにまとめる

- 分布モデルを仮定する＝モデルベースクラスタリング
  - 非階層的

    モデルが GMM

    ⋮
- **分布モデルを仮定しない**
  - 非階層的

    k-means 法
  - 階層的

    距離に基づく（重心間，ペア最短，Ward 法，…）

唯一の正解はない．評価指標は複数ある．
クラスター（成分）の個数 K を指定するもの，自動的に決まるものがある

データ空間の距離

データ点 $\boldsymbol{x}_i \in \mathbb{R}^p$.
x_{ij} $(i = 1, \dots, n, j = 1, \dots, p)$. i 番目のデータ (行) の j 番目のコラムの値.
近い $\Leftrightarrow$ 距離が小さい 距離空間, 位相
データ点 $\boldsymbol{x}_i, \boldsymbol{x}_{i'}$ の間の距離には様々なものが考えられる. 例.

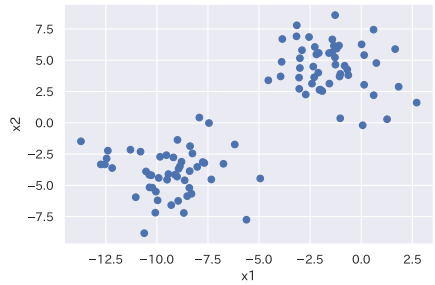
$$d(\boldsymbol{x}_i, \boldsymbol{x}_{i'}) = \left(\sum_{j=1}^{p} |x_{ij} - x_{i'j}|^r \right)^{1/r}$$

単位, 成分の意味, 分布によって, 距離の大小の基準は変わる. すべての j を平等に扱うには, 無難には, あらかじめデータを標準化しておく.

マハラノビスの距離

$$\tilde{x}_{ij} = \frac{x_{ij} - \overline{x}_j}{\sigma_j}$$

$$\tilde{x}_{ij} = \frac{x_{ij} - x_{\min\, j}}{x_{\max\, j} - x_{\min\, j}} \in [0, 1]$$



ここまで来たよ

9. 非/階層的クラスタリング
   - 様々なクラスタリングのアルゴリズム
   - （モデルを仮定しない）非階層的クラスタリング：K-means 法
   - 階層的クラスタリング, 距離に基づく, Ward 法
   - クラスタリングの評価
   - 対称行列の固有値・固有ベクトル・対角化

K-means のアルゴリズム

ハード，非階層的クラスタリング（K は途中や出力で一定）
ユーザが最初に，最終的クラスタの個数 K を指定する.

1. K 個の初期「クラスタの代表点」$\boldsymbol{x}_{\mathrm{C}k}$ $(k = 1, \ldots, K)$ を定める (バリエーションあり).
2. 収束するまで (すべてのデータ点の属するクラスタが変わらなくなるまで), 以下を繰り返す.
   1. n 個のデータ点に対して, 代表点 $\boldsymbol{x}_{\mathrm{C}k}$ が最も近いクラスタに属させてクラスタを更新する.
   2. $\boldsymbol{x}_{\mathrm{C}k}$ を, 更新されたクラスタに属するデータ点の平均 (mean) に更新する.

結果は初期「クラスタの代表点」に依存する
K をいくつにとるのがいいかを知るには, さまざまな K で試行して, 群内平方和などの指標で比較する
上では，$k = 1, \ldots, K$ という個数を用い，クラスタの個数は K だが，k-means というときは，クラスタの個数は小文字 k というつもり

L09-Q1

Quiz(非階層的クラスター分析 (k-means))

次の 2 変量データを考える.

	x_1	x_2
A	1	1
B	1	3
C	5	1
D	5	3

データ点間の距離はユークリッド距離, クラスターの代表位置は平均 (重心) とする.
k-means 法で $K = 2$ とする. 初期のクラスターの代表点を, $(1, 0), (3, 3)$ とする. 最初に得られるクラスターとその代表点, 更新後に得られるクラスターとその代表点, を求めよう.

略解

- $(1,0)$ を代表点とするクラスター 1: {A}, $(3,3)$ を代表位置とするクラスター 2: {B,C,D}.
- 更新後の代表点 クラスター 1: $(1,1)$, クラスター 2: $(\frac{1+5+5}{3}, \frac{3+1+1}{3})$
- クラスター 1: {A,B} クラスター 2: {C,D}. (B とクラスター 1 の距離 2, B とクラスター 2 との距離 2.75 なので.

ここまで来たよ

9. 非/階層的クラスタリング
   - 様々なクラスタリングのアルゴリズム
   - （モデルを仮定しない）非階層的クラスタリング：K-means 法
   - 階層的クラスタリング, 距離に基づく, Ward 法
   - クラスタリングの評価
   - 対称行列の固有値・固有ベクトル・対角化

（モデルを仮定しない）階層的クラスタリング

準備

データ点とデータ点の距離を,

- データ点とクラスタ (データ点の集合) の距離
- クラスターとクラスタの距離 (クラスター間距離)

に拡張する.

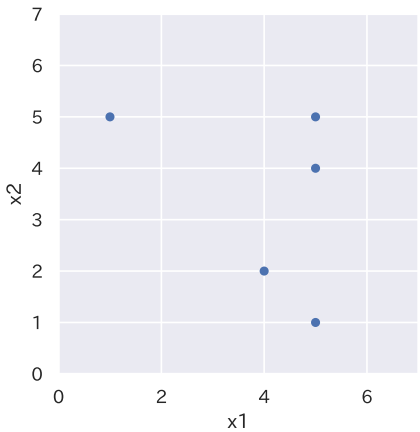
クラスター間の距離の様々な定義

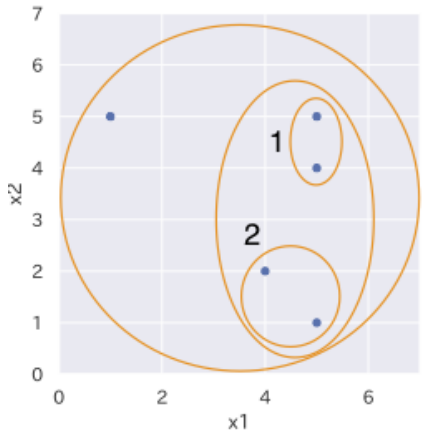
- クラスター内の平均 (重心) 間の距離
- (クラスター内の平均 (重心) 間の距離)$^2 \times (\frac{1}{2}(n_1^{-1} + n_2^{-1}))^{-1}$ (Ward 距離，Ward 法)
- データ点ペアの最短距離
- データ点ペアの最長距離
- データ点ペアの平均距離

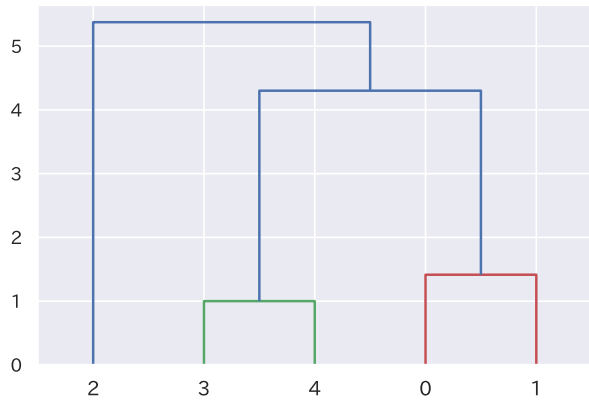
階層的クラスタリングのアルゴリズム

1. すべてのデータ点を, 1 点のみからなるクラスターとみなす. クラスターの個数 $K = n$.
2. 最短距離が指定の距離より大きい間, または K が指定の値より大きい間，次を反復する
   1. $K(K-1)/2$ ペアのクラスター間距離を計算する.
   2. 最短距離にある 2 クラスターを合併して 1 個のクラスターとみなす. K を 1 減らす.

合併の過程を描いたもの: デンドログラム (dendrogram, 樹形図)







- ハード，階層的クラスタリング
- 終了までの途中のどこかの K を人間が選んで採用する．あらかじめ，または後出しで．
- クラスタの分かれ方は,「望みのもの」になるとは限らない．各クラスタのサイズ（属するデータ点の個数）が同程度になる保証はない．
  - 1 つだけのクラスタが 1 点ずつ膨らんでいく鎖現象

L09-Q2

Quiz(階層的クラスター分析 (Ward 法))

次の 2 変量データを考える.

	x_1	x_2
A	10	2
B	10	4
C	2	10
D	10	8
E	10	10

距離を (ユークリッド距離)$^2\times$ (クラスターサイズの調和平均), クラスターの代表点は重心とする (Ward 法).
階層的クラスター分析 (Ward 法) を行い, デンドログラムを描こう.

略解

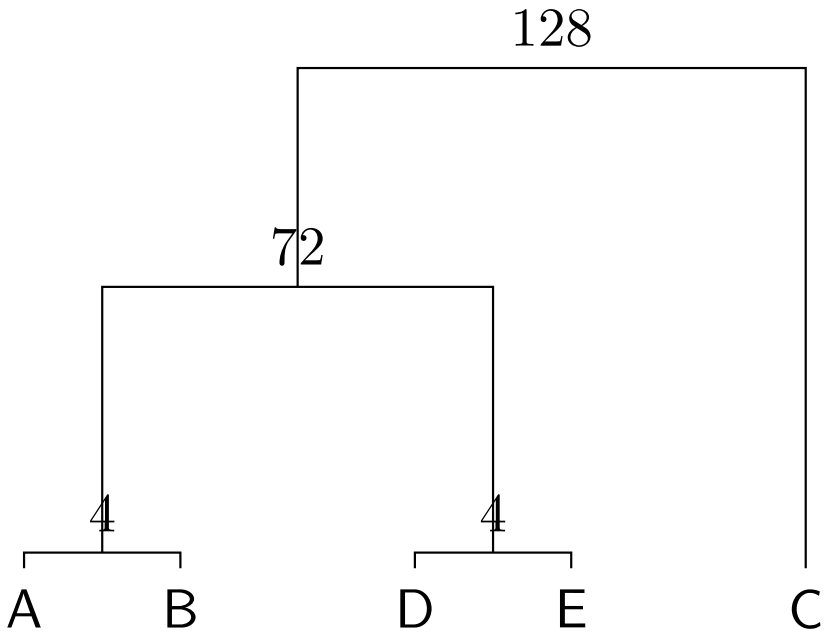
初期には，クラスターサイズはすべて 1 なので，ユークリッド距離だけで比べたのと大小は同じ.

AB, DE の距離はともに $(0^2+2^2)\cdot(\frac{1}{2}(1^{-1}+1^{-1}))^{-1}=4$ で最小だから，AB, DE がクラスターとなり，代表点 F,G は $(10,3)$ と $(10,9)$．合流の順序は逆でもよい．

$$d(F,G)=(0^2+6^2)\cdot(\tfrac{1}{2}(2^{-1}+2^{-1}))^{-1}=36\cdot 2=72,$$
$$d(F,C)=(8^2+7^2)\cdot(\tfrac{1}{2}(2^{-1}+1^{-1}))^{-1}=113\cdot\tfrac{4}{3}=\tfrac{452}{3},$$
$$d(G,C)=(8^2+1^2)\cdot(\tfrac{1}{2}(2^{-1}+1^{-1}))^{-1}=65\cdot\tfrac{4}{3}=\tfrac{260}{3}.$$

FG 間の距離が最小だから，F と G が先にクラスターにまとまる．代表点 H は $(10,6)$.
最後に，H と C が，$d(H,C)=(8^2+4^2)\cdot(\frac{1}{2}(4^{-1}+1^{-1}))^{-1}=80\cdot\frac{8}{5}=128$ で合流する．代表点は $\left(\frac{42}{5},\frac{34}{5}\right)$.

順序	合流	距離（デンドログラムの高さ）	合流後の代表点
1	A と B → F	4	$(10,3)$
2	D と E → G	4	$(10,9)$
3	F と G → H	72	$(10,6)$
4	H と C	128	$\left(\frac{42}{5},\frac{34}{5}\right)$


128
72
4
4
A
B
D
E
C

ここまで来たよ

9. 非/階層的クラスタリング
   - 様々なクラスタリングのアルゴリズム
   - （モデルを仮定しない）非階層的クラスタリング：K-means 法
   - 階層的クラスタリング, 距離に基づく, Ward 法
   - クラスタリングの評価
   - 対称行列の固有値・固有ベクトル・対角化

クラスタリングの様々な評価方法

手法	基準			
回帰	尤度 →	決定係数		
回帰	尤度+変数の個数 →	自由度調整済決定係数		
分類	尤度 →	決定係数 →	混同行列	
分類	尤度+閾値変更 →		→	ROC, AUC
クラスタリング	?			

これまで出てきた様々な手法の評価

基準		制約	考え方
凝集度 × 分離度	CH 基準		クラスタ内が近く，クラスタ間が遠ければよい
「正解率」	相互情報量	正解が必要	正解と似ていればよい
安定性	評価量の分散		初期値を変更して繰り返し，同じ結果になればよい
分布への適合	尤度，BIC, AIC	モデルベース	モデルにしたがう分布ならよい

平方和の分解による，凝集度，分離度の評価

本質は $p = 1$ 次元で見えるのでその表現で.
x_{ik}: k 番目のクラスターに属する i 番目のデータ点. $i = 1, \dots, n_k$. $k = 1, \dots, K$.
$\sum_{k=1}^{K} n_k = n$.
偏差平方和

$$
\begin{aligned}
S &= \sum_{i,k} [x_{ik} - \overline{x}_{..}]^2 \\
&= \sum_{i,k} [(x_{ik} - \overline{x}_{.k}) + (\overline{x}_{.k} - \overline{x}_{..})]^2 \\
&\overset{!}{=} \sum_{k} \sum_{i} (x_{ik} - \overline{x}_{.k})^2 + \sum_{k} n_k (\overline{x}_{.k} - \overline{x}_{..})^2 \\
&= S_{\mathrm{W}} + S_{\mathrm{B}} \\
&= \text{群内平方和} + \text{群間平方和}
\end{aligned}
$$

群=クラスター.
$x_{.k} = \frac{1}{n_k} \sum_i x_{ik}$ 群内平均
$x_{..} = \frac{1}{n} \sum_{i,k} x_{ik}$ 全体平均全平方和=回帰平方和+残差平方和 と似てる？

カリンスキ-ハラバシュ基準

定義 (カリンスキ-ハラバシュ (CH) 基準)

$$\mathrm{CH} = \frac{(n-K)S_{\mathrm{B}}}{(K-1)S_{\mathrm{W}}}$$

- CH が 大きいほどよいクラスタリング.
- クラスタの個数 K が異なる場合にも対応させた（回帰の自由度調整済決定係数みたいに）
- 実は, LDA(線形判別分析) の Fisher の線形判別関数は比 $S_{\mathrm{B}}/S_{\mathrm{W}}$ を最大化するものになっていた.
- Ward 法=CH を最大化するクラスタリング．クラスター間距離=(そのクラスターを合併したときの群内平方和 S_{W} の増分)，とした時の，距離に基づくクラスタリング

ここまで来たよ

9. 非/階層的クラスタリング
   - 様々なクラスタリングのアルゴリズム
   - （モデルを仮定しない）非階層的クラスタリング：K-means 法
   - 階層的クラスタリング, 距離に基づく, Ward 法
   - クラスタリングの評価
   - 対称行列の固有値・固有ベクトル・対角化

行列の固有値・固有ベクトル

定義 (行列の固有値・固有ベクトル 石村畑 §6.3(p.214))

A: n 次実正方行列
ある $\lambda \in \mathbb{C}$ に対し，

$$A\boldsymbol{v} = \lambda\boldsymbol{v}$$

を満たす $\boldsymbol{v} \in \mathbb{R}^n$（ただし，$\boldsymbol{v} \neq 0$) が存在するとき，
λ を A の固有値,
$\boldsymbol{v}$ を固有値 λ に属する固有ベクトルという

定理 (固有方程式 石村畑 定理 6.3.1(p.214))

n 次正方行列 A に対し，

$$\lambda \in A\text{は } A \text{ の固有値} \iff \lambda\text{は固有方程式} \det(\lambda E - A) = 0\text{の解}$$

定義 (対角化 石村畑 §6.4(p.220))

n 次正方行列 A に対して，正則行列 P を見つけて $P^{-1}AP = D$ を対角行列 石村畑 §6.4 にすることを対角化という．

対称行列の固有値固有ベクトル

定義 (対称行列 石村畑 §6.5(p.234))

実行列 A が対称行列 $\Leftrightarrow A^\top = A$.

定義 (対称行列の固有値固有ベクトル対角化 石村畑 定理 6.5.1,6.5.2,6.5.3)

- 対称行列の固有値は実数
- 対称行列の相異なる固有値に属する固有ベクトルは直交
- 対称行列は直交行列で対角化可能

対称行列の対角化の手順 石村畑 §6.5(p235)

1. 固有方程式 $\det(\lambda E - A) = 0$ を解いて，固有値 $\lambda_1, \dots, n$ を求める．
2. 各 λ_i に属する固有ベクトル $\boldsymbol{v}_i$ を，1 次方程式 $A\boldsymbol{v}_i = \lambda \boldsymbol{v}_i$ を解いて求める．
3. シュミットの方法 石村畑 §6.1 で正規直交化した $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ を取り直す．
4. 直交行列 石村畑 §6.2 $V = (\boldsymbol{v}_1, \dots, \boldsymbol{v}_n)$ により，A は次のように対角化される．

$$V^{-1}AV = V^{\top}AV = D = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}, \quad A = VDV^{\top}$$

定義 (直交行列 石村畑 §6.2)

$M^{\top} = M^{-1}$ を満たす行列を直交行列という．

M が直交行列である $\Longleftrightarrow M$ の列ベクトルが正規直交系をなす．

L09-Q3

Quiz(実対称行列の直交行列による対角化)

実対称行列

$$A = \begin{pmatrix} \frac{23}{5} & \frac{36}{5} \\ \frac{36}{5} & \frac{2}{5} \end{pmatrix}$$

の固有値固有ベクトルを求めよう.

略解

固有方程式 $\det(A - \lambda E) = 0$ を解いて, $\lambda_1 = 10, \lambda_2 = -5$.
固有値 $\lambda_1 = 10$ に対応する固有ベクトルは $\boldsymbol{u}_1 = \begin{pmatrix} 4 \\ 3 \end{pmatrix} s$ $(s \in \mathbb{R}, s \neq 0)$, 固有値 $\lambda_2 = -5$ に対応する固有ベクトルは $\boldsymbol{u}_2 = \begin{pmatrix} -3 \\ 4 \end{pmatrix}$ $(s \in \mathbb{R}, s \neq 0)$.

L09-Q4

Quiz(実対称行列の直交行列による対角化)

実対称行列

$$A = \begin{pmatrix} \frac{23}{5} & \frac{36}{5} \\ \frac{36}{5} & \frac{2}{5} \end{pmatrix}$$

を, 直交行列で対角化しよう.

略解

規格化された ($|\boldsymbol{u}| = 1$ である) 固有ベクトルから正則行列 P を作ると,
正則行列 P は直交行列となり, $P^{-1} = P^{\top}$ である.
固有値 $\lambda_1 = 10$ に対応する規格化された固有ベクトルのひとつは
$\boldsymbol{v}_1 = \frac{1}{5}\begin{pmatrix} 4 \\ 3 \end{pmatrix}$.
固有値 $\lambda_2 = -5$ に対応する規格化された固有ベクトルのひとつは
$\boldsymbol{v}_2 = \frac{1}{5}\begin{pmatrix} -3 \\ 4 \end{pmatrix}$.

よって, 直交行列 $P = \begin{pmatrix} \frac{4}{5} & -\frac{3}{5} \\ \frac{3}{5} & \frac{4}{5} \end{pmatrix}$ で,

$$P^\top A P = \Lambda$$

と対角化される. ただし, 対角行列 $\Lambda = \left(\begin{smallmatrix} 10 & 0 \\ 0 & -5 \end{smallmatrix} \right)$.