

問題解決技法入門

Cluster analysis using R

堀田 敬介

Rでクラスター分析

- Rを起動, csv ファイルをデータとして読み込み
 - 「マイドキュメント(Y:)」の「R」フォルダに保存

data-seiseki.csv

	算数	理科	国語	英語	社会
太郎	90	100	70	90	30
次郎	80	60	70	70	20
三郎	100	40	30	70	80
四郎	60	30	40	80	80
花子	30	60	80	90	90
寒子	50	60	40	30	60
湘子	90	100	90	80	70

※ファイルのフルパス

例) YドライブのRフォルダ内にあるdata-seiseki.csvという名前のファイル

• R Console

※1行目にheaderあり ※各行1列目は名前

```
> seiseki <- read.csv("Y:/R/data-seiseki.csv", header=T, row.names=1)
```

※csvファイルを読み込み, 変数seiseki に代入

Rでクラスター分析

- 関数dist()で距離を計算し, seiseki.dに代入

```
> seiseki.d <- dist(seiseki, "manhattan")
```

※マンハッタン距離("manhattan")を用いて距離を計算している
他の距離を使いたいときは
"manhattan"を右に変更

"euclidean" = ユークリッド距離
"minkowski", p=3 = p=3のミンコフスキー距離
"maximum" = l_{∞} ノルム (える むげんだい のるむ)

- 階層クラスター分析をし, 結果をseiseki.hcに代入

```
> (seiseki.hc <- hclust(seiseki.d, "ward"))
```

※ワード法("ward")を用いてクラスター分析を実施している
他の方法を使いたいときは、"ward"を以下に変更

"single" = 最短距離法,	"complete" = 最長距離法	
"average" = 群平均法,	"centroid" = 重心法,	"median" = 中央値法

- 結果をデンドログラム(樹形図)で描画

```
> plot(seiseki.hc, hang=-1)
```