

演習 クラスター分析

クラスター分析の目的

1) 類似度による個体（レコード）の分類

2) 類似度による変数の分類

クラスター分析は分類をどのように表示するか → デンドrogram

デンドrogramの縦軸は → 要素またはクラスター間の距離（類似の程度を示す量）

要素間の距離とは

個体間について

量的データ：ユークリッド距離、標準化ユークリッド距離、マハラノビス距離等

質的0/1データ：類似比、一致係数、 ϕ 係数等を使ったもの

変数間について

量的データ：相関係数、順位相関係数等を使ったもの

質的データ：平均平方根一致係数、一致係数、クラメールのV等を使ったもの

要素間の距離を知るには → 距離行列

クラスター構成法

最短距離法（棒状の分布に最適）

最長距離法（クラスターを分離する能力が高い）

他に、群平均法、重心法、メジアン法、ウォード法

クラスター構成過程を表示するには → クラスター構成と距離

演習1

多変量演習9.txtは学生による授業評価のデータであり、レコード（個体）は1つの授業で調べた質問項目（変数）ごとの平均を表している。このデータからクラスター分析を用いて、個体や変数の類似性の特徴を見出したい。以下の質問に答えよ。

1) ユークリッド距離を用いた場合、1番と12番の距離はいくらか。[]

2) クラスター構成法を最長距離法、距離測定法をユークリッド距離とする場合、最初にクラスターを構成するのは何番と何番でそれらの距離はいくらか。

個体 [] 番と個体 [] 番で、距離 []

3) 上の設定で、最初にクラスターとクラスター、またはクラスターと要素の結合になるのはどのようなクラスター（要素）か。それらに含まれる要素を示せ。またその際の距離はいくらか。

クラスター [] とクラスター（要素）[] 距離 []

4) 上の設定でクラスター分析を実行し、4つのクラスターに分けたとき、それらのクラスターに含まれる要素（授業の番号）は何か。

[] [] [] []

5) 5番が含まれるクラスターと10番が含まれるクラスターの最も大きな特徴は何か。5番 [] 10番 []

6) 距離測定法を標準化ユークリッド距離（各変数を標準化したときのユークリッド距離）に変えた場合、クラスター構成は大きく変わるか。

[変わる・あまり変わらない] 注) 標準化値 = (値 - 平均値) / 標準偏差

7) これにはどんな理由が考えられるか。

各変数の [] があまり違わないから。

8) 距離測定法をユークリッド距離とし、クラスター構成法を最短距離法に変えるとクラスター構成は大きく変わるか。[変わる・あまり変わらない]

9) ユークリッド距離の場合、その他のクラスター構成法は最長距離法と最短距離法のどちらに近いか。[最長距離法・最短距離法]

各質問についての分類を行いたいが、距離測定法を1-相関係数として以下の問い合わせに答えよ。

10) 最長距離法で上の距離測定法を用いる場合、最初にクラスターを構成するのは何と何で、そのときの距離はいくらか。

変数 [] と変数 [] で、距離 []

11) 上の設定でクラスター分析を行い、変数を3つのクラスターに分類する場合、それらのクラスターに含まれる要素（変数）は何か。

[] [] [] []

演習解答（多変量演習 9.txt）

1) ユークリッド距離を用いた場合、1番と12番の距離はいくらか。[0.404]

2) クラスター構成法を最長距離法、距離測定法をユークリッド距離とする場合、最初にクラスターを構成するのは何番と何番でそれらの距離はいくらか。

個体 [15] 番と個体 [16] 番で、距離 [0.199]

3) 上の設定で、最初にクラスターとクラスター、またはクラスターと要素の結合になるのはどのようなクラスター（要素）か。それらに含まれる要素を示せ。またその際の距離はいくらか。

クラスター [8,14] とクラスター（要素）[15,16] 距離 [0.568]

4) 上の設定でクラスター分析を実行し、4つのクラスターに分けたとき、それらのクラスターに含まれる要素（授業の番号）は何か。

[1,12,4,3,18,2,20] [8,14,15,16,17] [9,10,19] [5,11,13,6,7]

5) 5番が含まれるクラスターと10番が含まれるクラスターの最も大きな特徴は何か。

5番 [高い評価] 10番 [低い評価]

6) 距離測定法を標準化ユークリッド距離（各変数を標準化したときのユークリッド距離）に変えた場合、クラスター構成は大きく変わるか。

[変わる・あまり変わらない] 注) 標準化値 = (値 - 平均値) / 標準偏差

7) これにはどんな理由が考えられるか。

各変数の [標準偏差] があまり違わないから。

8) 距離測定法をユークリッド距離とし、クラスター構成法を最短距離法に変えるとクラスター構成は大きく変わるか。[変わる・あまり変わらない]

9) ユークリッド距離の場合、その他のクラスター構成法は最長距離法と最短距離法のどちらに近いか。[最長距離法・最短距離法]

10) 最長距離法で上の距離測定法を用いる場合、最初にクラスターを構成するのは何と何で、そのときの距離はいくらか。

変数 [分かり易さ] と変数 [有益さ] で、距離 [0.100]

11) 上の設定でクラスター分析を行い、変数を3つのクラスターに分類する場合、それらのクラスターに含まれる要素（変数）は何か。

[進む速さ,黒板等,分かり易さ,有益さ] [声の大きさ,受講態度] [死後注意]