

データサイエンスの諸相（多次元尺度法）

前田利之（まえだ としゆき）

maechan@hannan-u.ac.jp

阪南大学 経営情報学部

(2024.11.06)

多次元尺度法

多次元尺度法

() とは

- 要素ごとの近さ（距離）が定義されている前提
- 多次元データを のグラフとして表示する方法
- データは常に数量化されているとは限らない
- → を考え、その上で を考える

多次元尺度法

多次元尺度法

(**multidimensional scaling, MDS**) とは

- 要素ごとの近さ（距離）が定義されている前提
- 多次元データを のグラフとして表示する方法
- データは常に数量化されているとは限らない
- → を考え、その上で を考える

多次元尺度法

多次元尺度法

(**multidimensional scaling, MDS**) とは

- 要素ごとの近さ（距離）が定義されている前提
- 多次元データを **低(2,3)次元** のグラフとして表示する方法
- データは常に数量化されているとは限らない
- → を考え、その上で を考える

多次元尺度法

多次元尺度法

(**multidimensional scaling, MDS**) とは

- 要素ごとの近さ（距離）が定義されている前提
- 多次元データを **低(2,3)次元** のグラフとして表示する方法
- データは常に数量化されているとは限らない
- → **尺度** を考え、その上で を考える

多次元尺度法

多次元尺度法

(**multidimensional scaling, MDS**) とは

- 要素ごとの近さ（距離）が定義されている前提
- 多次元データを **低(2,3)次元** のグラフとして表示する方法
- データは常に数量化されているとは限らない
- → **尺度** を考え、その上で **距離** を考える

多次元とは？

- 1次元：
- 2次元：
- 3次元：
- ... ここまではイメージできる（!？）

多次元とは？

- 1次元： 直線上
- 2次元：
- 3次元：
- ... ここまではイメージできる（!？）

多次元とは？

- 1次元： 直線上
- 2次元： 平面上
- 3次元：
- ... ここまではイメージできる（!？）

多次元とは？

- 1次元： 直線上
- 2次元： 平面上
- 3次元： 立体空間
- ... ここまではイメージできる（!？）

多次元とは？

- 4、5、6、…次元：イメージはつきにくいが、物理量としては考えられる
 - たとえば5次元ベクトル $(3, 2, 1, 1, 1)$, $(1, 2, 3, 1, 0)$ の間の距離は？…

…計算可能

多次元とは？

- 4、5、6、…次元：イメージはつきにくいですが、物理量としては考えられる
 - たとえば5次元ベクトル $(3, 2, 1, 1, 1)$, $(1, 2, 3, 1, 0)$ の間の距離は？…

$$\sqrt{(1-3)^2 + (2-2)^2 + (3-1)^2 + (1-1)^2 + (0-1)^2}$$

…計算可能

多次元とは？

- 4、5、6、…次元：イメージはつきにくいですが、物理量としては考えられる
 - たとえば5次元ベクトル $(3, 2, 1, 1, 1)$, $(1, 2, 3, 1, 0)$ の間の距離は？…

$$\sqrt{(1-3)^2 + (2-2)^2 + (3-1)^2 + (1-1)^2 + (0-1)^2}$$

$$= \sqrt{4 + 0 + 4 + 0 + 1} = 3 \cdots \text{計算可能}$$

尺度

- そもそもデータマイニングとは、データの中にある特徴を抽出すること
- そのときにはデータを として表し、数理的
手法をほどこす
 - 値自体を足したり引いたりできるデータ =
 - 名前など、値をもたないデータ =
 - 数字で表していても **量を表しているとは限らない**

尺度

- そもそもデータマイニングとは、データの中にある特徴を抽出すること
- そのときにはデータを **数量** として表し、数理的手法をほどこす
 - 値自体を足したり引いたりできるデータ =
 - 名前など、値をもたないデータ =
 - 数字で表していても **量を表しているとは限らない**

尺度

- そもそもデータマイニングとは、データの中にある特徴を抽出すること
- そのときにはデータを **数量** として表し、数理的
手法をほどこす
 - 値自体を足したり引いたりできるデータ＝
量的データ
 - 名前など、値をもたないデータ＝
 - 数字で表していても **量を表しているとは限らない**

尺度

- そもそもデータマイニングとは、データの中にある特徴を抽出すること
- そのときにはデータを **数量** として表し、数理的手法をほどこす
 - 値自体を足したり引いたりできるデータ = **量的データ**
 - 名前など、値をもたないデータ = **質的データ**
 - 数字で表していても **量を表しているとは限らない**

尺度（続き）

- :扱うデータを区別するために割り当てた番号（学籍番号、等）
- :何かを基準に並べた順番（行列待ちの番号、等）
- :それぞれの差が意味があるもの（温度、身長、等）
- :間隔だけでなく、比が意味をもつもの

尺度（続き）

- **名義尺度**:扱うデータを区別するために割り当てた番号（学籍番号、等）
- :何かを基準に並べた順番（行列待ちの番号、等）
- :それぞれの差が意味があるもの（温度、身長、等）
- :間隔だけでなく、比が意味をもつもの

尺度（続き）

- **名義尺度**:扱うデータを区別するために割り当てた番号（学籍番号、等）
- **順序尺度**:何かを基準に並べた順番（行列待ちの番号、等）
- :それぞれの差が意味があるもの（温度、身長、等）
- :間隔だけでなく、比が意味をもつもの

尺度（続き）

- **名義尺度**:扱うデータを区別するために割り当てた番号（学籍番号、等）
- **順序尺度**:何かを基準に並べた順番（行列待ちの番号、等）
- **間隔尺度**:それぞれの差が意味があるもの（温度、身長、等）
- :間隔だけでなく、比が意味をもつもの

尺度（続き）

- **名義尺度**:扱うデータを区別するために割り当てた番号（学籍番号、等）
- **順序尺度**:何かを基準に並べた順番（行列待ちの番号、等）
- **間隔尺度**:それぞれの差が意味があるもの（温度、身長、等）
- **比例尺度**:間隔だけでなく、比が意味をもつもの

尺度（続き）

■ 名義尺度、順序尺度は

■ 間隔尺度、比例尺度は

■ 順序尺度と間隔尺度の違い

- 例えば身長が 195cm, 185cm, 140cm, 125cm, 110cm の 5 人がいて、高い順に 1 番、2 番と番号をふる
- 1 番と 2 番の身長差と 2 番と 3 番の身長差は異なる（以下同様）＝順序以外に意味はない

尺度（続き）

- 名義尺度、順序尺度は

質的データ

- 間隔尺度、比例尺度は

- 順序尺度と間隔尺度の違い

- 例えば身長が 195cm, 185cm, 140cm, 125cm, 110cm の 5 人がいて、高い順に 1 番、2 番と番号をふる
- 1 番と 2 番の身長差と 2 番と 3 番の身長差は異なる（以下同様）＝順序以外に意味はない

尺度（続き）

- 名義尺度、順序尺度は 質的データ
- 間隔尺度、比例尺度は 量的データ
- 順序尺度と間隔尺度の違い
 - 例えば身長が 195cm, 185cm, 140cm, 125cm, 110cm の 5 人がいて、高い順に 1 番、2 番と番号をふる
 - 1 番と 2 番の身長差と 2 番と 3 番の身長差は異なる（以下同様）＝順序以外に意味はない

尺度（続き）

- 順序尺度を間隔尺度として扱う場合もあるが、その場合は間隔がどうなっているのかを考慮する必要あり

- アンケートで「好き」～「嫌い」を5段階で扱う場合、それらに5～1を割り振り、平均をとったりする（実際は「好き」「少し好き」「どちらでもない」の差は等しいとは限らない）
- 前の例：身長が195cm, 185cm, 140cm, 125cm, 110cm

の5人を表にすると：

区分	人数
110cm ~ 129cm	2
130cm ~ 149cm	1
150cm 以上	2

…印象としては130と150の間に平均がありそうだが、実際は

尺度（続き）

- 順序尺度を間隔尺度として扱う場合もあるが、その場合は間隔がどうなっているのかを考慮する必要あり

- アンケートで「好き」～「嫌い」を5段階で扱う場合、それらに5～1を割り振り、平均をとったりする（実際は「好き」「少し好き」「どちらでもない」の差は等しいとは限らない）
- 前の例：身長が195cm, 185cm, 140cm, 125cm, 110cm

の5人を表にすると：

区分	人数
110cm ~ 129cm	2
130cm ~ 149cm	1
150cm 以上	2

…印象としては130と150の間に平均がありそうだが、実際は **151**

尺度（続き）

■ 間隔尺度と比例尺度の違い

- 例えば 1000 円は 100 円の 10 倍…

- （摂氏）温度の場合、 30°C は 10°C の 3 倍暑いわけ
はない…

- 絶対温度 (からの温度) なら意味があるかも

- ⇒ 基準である 0（原点）に意味がある場合＝

- データマイニングの際には本来の尺度でなくても間隔尺度や比例尺度と仮定する場合があるが、そのときにはその**仮定をしっかりと考慮しなくてはならない！**

尺度（続き）

■ 間隔尺度と比例尺度の違い

- 例えば 1000 円は 100 円の 10 倍…

比が意味をもつ

- （摂氏）温度の場合、 30°C は 10°C の 3 倍暑いわけではない…

- 絶対温度 (からの温度) なら意味があるかも

- ⇒ 基準である 0（原点）に意味がある場合＝

- データマイニングの際には本来の尺度でなくても間隔尺度や比例尺度と仮定する場合があるが、そのときにはその**仮定をしっかりと考慮しなくてはならない！**

尺度（続き）

■ 間隔尺度と比例尺度の違い

- 例えば 1000 円は 100 円の 10 倍…

比が意味をもつ

- （摂氏）温度の場合、 30°C は 10°C の 3 倍暑いわけではない… 比が意味をもたない

- 絶対温度 ()
からの温度) なら意味があるかも

- ⇒ 基準である 0（原点）に意味がある場合＝

- データマイニングの際には本来の尺度でなくても間隔尺度や比例尺度と仮定する場合があるが、そのときにはその**仮定をしっかりと考慮しなくてはならない！**

尺度（続き）

■ 間隔尺度と比例尺度の違い

- 例えば 1000 円は 100 円の 10 倍…

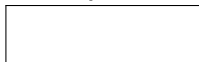
比が意味をもつ

- （摂氏）温度の場合、 30°C は 10°C の 3 倍暑いわけでは

ない… 比が意味をもたない

- 絶対温度 (-273.15°C からの温度) なら意味があるかも

- ⇒ 基準である 0（原点）に意味がある場合＝



- データマイニングの際には本来の尺度でなくても間隔尺度や比例尺度と仮定する場合があるが、そのときにはその**仮定をしっかりと考慮しなくてはならない！**

尺度（続き）

■ 間隔尺度と比例尺度の違い

- 例えば 1000 円は 100 円の 10 倍…

比が意味をもつ

- （摂氏）温度の場合、 30°C は 10°C の 3 倍暑いわけ

ではない… 比が意味をもたない

- 絶対温度 (-273.15°C) からの温度) なら意味があるかも

- ⇒ 基準である 0（原点）に意味がある場合＝

比例尺度

- データマイニングの際には本来の尺度でなくても間隔尺度や比例尺度と仮定する場合があるが、そのときにはその**仮定をしっかりと考慮しなくてはならない！**

距離

■ 直観的な 2 点間の距離 = 距離

■ 実は、 を満たせばほかの定義でもよい: 2 点 p, q とその間の関数 $d(p, q)$ について

1 : $d(p, q) \geq 0$ … 距離は負はならない

2 : $p = q \Leftrightarrow d(p, q) = 0$ … 距離が 0なのは
2 点が同じときだけ (それ以外は正)

3 : $d(p, q) = d(q, p)$

4 : 3 点 p, q, r にたいして
 $d(p, q) \leq d(p, r) + d(r, q)$

距離

■ 直観的な 2 点間の距離 = ユークリッド距離

■ 実は、 を満たせばほかの定義でもよい: 2 点 p, q とその間の関数 $d(p, q)$ について

1 : $d(p, q) \geq 0$ …距離は負はならない

2 : $p = q \Leftrightarrow d(p, q) = 0$ …距離が 0なのは 2 点が同じときだけ (それ以外は正)

3 : $d(p, q) = d(q, p)$

4 : 3 点 p, q, r にたいして
 $d(p, q) \leq d(p, r) + d(r, q)$

距離

■ 直観的な 2 点間の距離 = ユークリッド 距離

■ 実は、距離の公理を満たせばほかの定義でもよい: 2 点 p, q とその間の関数 $d(p, q)$ について

1 : $d(p, q) \geq 0$... 距離は負はならない

2 : $p = q \Leftrightarrow d(p, q) = 0$... 距離が 0なのは
2 点が同じときだけ (それ以外は正)

3 : $d(p, q) = d(q, p)$

4 : 3 点 p, q, r にたいして
 $d(p, q) \leq d(p, r) + d(r, q)$

距離

- 直観的な 2 点間の距離 = ユークリッド距離
- 実は、距離の公理を満たせばほかの定義でもよい: 2 点 p, q とその間の関数 $d(p, q)$ について
 - 1 非負性: $d(p, q) \geq 0$ … 距離は負はならない
 - 2 : $p = q \Leftrightarrow d(p, q) = 0$ … 距離が 0なのは 2 点と同じときだけ (それ以外は正)
 - 3 : $d(p, q) = d(q, p)$
 - 4 : 3 点 p, q, r にたいして $d(p, q) \leq d(p, r) + d(r, q)$

距離

- 直観的な2点間の距離＝ **ユークリッド** 距離
- 実は、**距離の公理** を満たせばほかの定義でもよい: 2点 p, q とその間の関数 $d(p, q)$ について
 - 1 **非負性**: $d(p, q) \geq 0$ … 距離は負はならない
 - 2 **非退化性**: $p = q \Leftrightarrow d(p, q) = 0$ … 距離が0なのは2点が同じときだけ (それ以外は正)
 - 3 : $d(p, q) = d(q, p)$
 - 4 : 3点 p, q, r にたいして
 $d(p, q) \leq d(p, r) + d(r, q)$

距離

- 直観的な 2 点間の距離 = ユークリッド 距離
- 実は、距離の公理 を満たせばほかの定義でもよい: 2 点 p, q とその間の関数 $d(p, q)$ について
 - 1 非負性 : $d(p, q) \geq 0$ … 距離は負はならない
 - 2 非退化性 : $p = q \Leftrightarrow d(p, q) = 0$ … 距離が 0なのは 2 点が同じときだけ (それ以外は正)
 - 3 対称性 : $d(p, q) = d(q, p)$
 - 4 : 3 点 p, q, r にたいして
$$d(p, q) \leq d(p, r) + d(r, q)$$

距離

- 直観的な 2 点間の距離 = ユークリッド距離
- 実は、距離の公理を満たせばほかの定義でもよい: 2 点 p, q とその間の関数 $d(p, q)$ について
 - 1 非負性: $d(p, q) \geq 0$ … 距離は負はならない
 - 2 非退化性: $p = q \Leftrightarrow d(p, q) = 0$ … 距離が 0なのは 2 点が同じときだけ (それ以外は正)
 - 3 対称性: $d(p, q) = d(q, p)$
 - 4 三角不等式: 3 点 p, q, r にたいして
$$d(p, q) \leq d(p, r) + d(r, q)$$

距離（続き）

- 対称性: $d(p, q) = d(q, p)$ …あたりまえのようであるが、実は距離の定義によっては等しくならない（例：飛行機での移動時間を考えると、多くの場合)

- 三角不等式: 3点 p, q, r にたいして $d(p, q) \leq d(p, r) + d(r, q)$ の意味は？

- 点 p から q への距離とは

- 途中に点 r を通った場合、 r がたまたま通り道にある場合を除いて になる、ということ…つまり

距離（続き）

- 対称性: $d(p, q) = d(q, p)$ …あたりまえのようであるが、実は距離の定義によっては等しくならない（例：飛行機での移動時間を考えると、多くの場合 **西向き $\geq$ 東向き**）

- 三角不等式: 3点 p, q, r にたいして $d(p, q) \leq d(p, r) + d(r, q)$ の意味は？

- 点 p から q への距離とは

- 途中に点 r を通った場合、 r がたまたま通り道にある場合を除いて になる、ということ…つまり

距離（続き）

- 対称性: $d(p, q) = d(q, p)$ …あたりまえのようであるが、実は距離の定義によっては等しくならない（例：飛行機での移動時間を考えると、多くの場合 **西向き $\geq$ 東向き**）

- 三角不等式: 3点 p, q, r にたいして $d(p, q) \leq d(p, r) + d(r, q)$ の意味は？

- 点 p から q への距離とは **最短距離**
- 途中に点 r を通った場合、 r がたまたま通り道にある場合を除いて になる、ということ…つまり

距離（続き）

- 対称性: $d(p, q) = d(q, p)$ …あたりまえのようであるが、実は距離の定義によっては等しくならない（例：飛行機での移動時間を考えると、多くの場合 **西向き $\geq$ 東向き**）
- 三角不等式: 3点 p, q, r にたいして $d(p, q) \leq d(p, r) + d(r, q)$ の意味は？
 - 点 p から q への距離とは **最短距離**
 - 途中に点 r を通った場合、 r がたまたま通り道にある場合を除いて **遠回り** になる、ということ…つまり



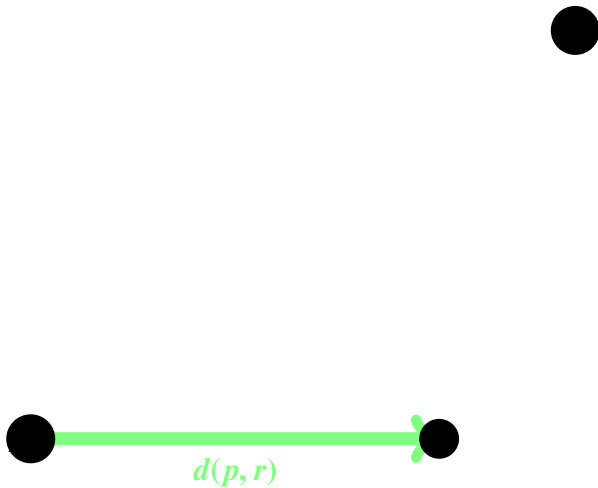
距離（続き）

- 対称性: $d(p, q) = d(q, p)$ …あたりまえのようであるが、実は距離の定義によっては等しくならない（例：飛行機での移動時間を考えると、多くの場合 **西向き $\geq$ 東向き**）
- 三角不等式: 3点 p, q, r にたいして $d(p, q) \leq d(p, r) + d(r, q)$ の意味は？
 - 点 p から q への距離とは **最短距離**
 - 途中に点 r を通った場合、 r がたまたま通り道にある場合を除いて **遠回り** になる、ということ…つまり **ベクトルの和**

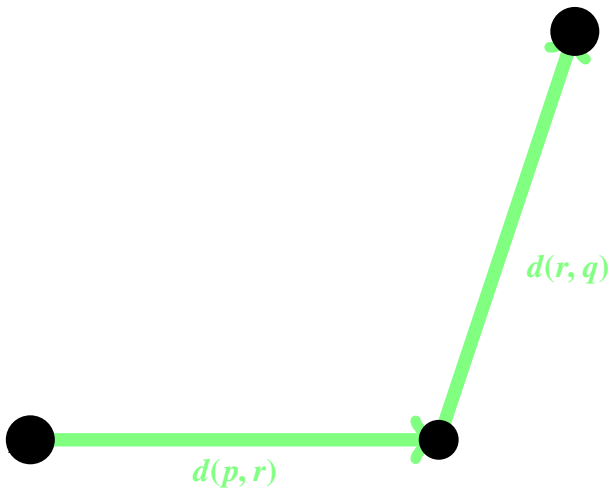
ベクトルの和



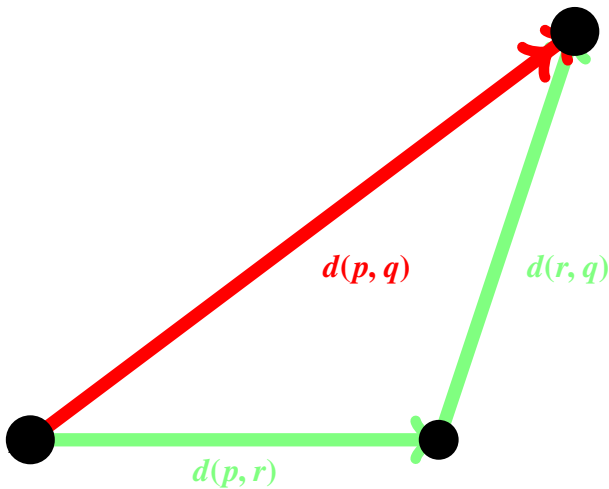
ベクトルの和



ベクトルの和



ベクトルの和



ベクトルの距離

☆ 【練習】 Rで

```
aa <- c(3, 2, 1, 1, 1)
```

```
bb <- c(1, 2, 3, 1, 0)
```

として、 $\vec{bb} - \vec{aa}$ の大きさを計算して値を確認しましょう。Rで $\vec{x}, \vec{y}$ の差は以下の通りです（既出）。

```
sqrt( sum( (x-y) * (x-y) ) )
```

☆ 差の長さ = ベクトルがさす 2点間の距離

多次元尺度法

- 数量化されたデータに距離が定義されているものとする
- 座標は与えられていない and/or 座標が多次元でグラフ化できない
- ⇒ 距離情報から で表す手法＝
- 可視化のため の座標で表現する

多次元尺度法

- 数量化されたデータに距離が定義されているものとする
- 座標は与えられていない and/or 座標が多次元でグラフ化できない
- ⇒ 距離情報から **空間内の点** で表す手法＝

- 可視化のため  の座標で表現する

多次元尺度法

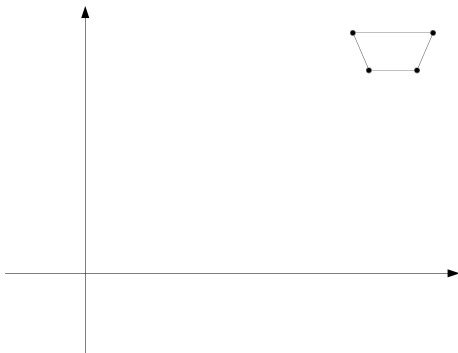
- 数量化されたデータに距離が定義されているものとする
- 座標は与えられていない and/or 座標が多次元でグラフ化できない
- ⇒ 距離情報から **空間内の点** で表す手法＝
多次元尺度法
- 可視化のため の座標で表現する

多次元尺度法

- 数量化されたデータに距離が定義されているものとする
- 座標は与えられていない and/or 座標が多次元でグラフ化できない
- ⇒ 距離情報から **空間内の点** で表す手法＝
多次元尺度法
- 可視化のため **低 (2 or 3) 次元** の座標で表現する

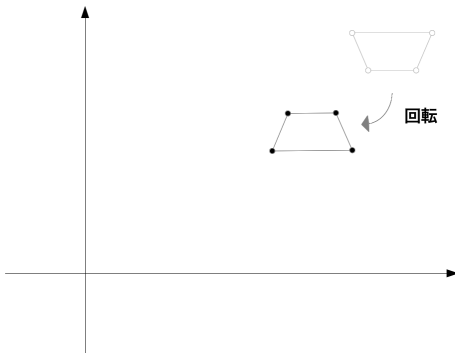
多次元尺度法のイメージ

☆ 回転と平行移動により各距離は変更せず変換処理



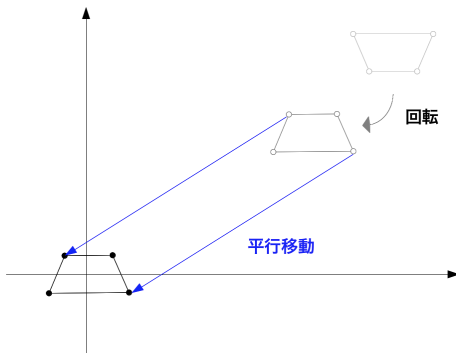
多次元尺度法のイメージ

☆ 回転と平行移動により各距離は変更せず変換処理



多次元尺度法のイメージ

☆ 回転と平行移動により各距離は変更せず変換処理



多次元尺度法（続き）

☆ 以下、式の厳密な展開はさておき、大まかな流れをつかむことが大切です…

- 1 問題の定式化: r 次元の空間における N 個の点それぞれに 2 点間の距離が与えられているとする。その座標を表す行列は以下のようなになる

$$X = \left(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N \right)^T = \begin{pmatrix} x_{11} & \dots & x_{1r} \\ \dots & \ddots & \dots \\ x_{N1} & \dots & x_{Nr} \end{pmatrix}$$

多次元尺度法（続き）

- 2 距離 d_{ij} をユークリッド距離とすると、各点の座標を表すベクトル $\vec{x}_i, \vec{x}_j$ を用いて

$$\begin{aligned} d_{ij}^2 &= (x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots (x_{ir} - x_{jr})^2 \\ &= \sum_{k=1}^r (x_{ik} - x_{jk})^2 = \vec{x}_i \vec{x}_i + \vec{x}_j \vec{x}_j - 2\vec{x}_i \vec{x}_j \end{aligned}$$

多次元尺度法（続き）

- 3 d_{ij}^2 を成分とする、次のような距離行列を $D^{(2)}$ とする

$$D^{(2)} = \begin{pmatrix} 0 & d_{12}^2 & d_{13}^2 & \dots & d_{1N}^2 \\ d_{21}^2 & 0 & d_{23}^2 & \dots & d_{2N}^2 \\ \dots & \dots & \dots & \dots & \dots \\ d_{N1}^2 & d_{N2}^2 & \dots & \dots & 0 \end{pmatrix}$$

- 4 この $D^{(2)}$ の値を前提に X を求めることを考える

多次元尺度法（続き）

5 $D^{(2)} = \text{diag}(XX^T) \cdot (\vec{1} \cdot \vec{1}^T) - 2XX^T$
 $+ (\vec{1} \cdot \vec{1}^T) \cdot \text{diag}(XX^T)$

を $D^{(2)}$ は満たす。ただし diag は対角成分のみ取りだし他は0とする関数であり、 $\vec{1}$ は次のベクトル

である。 $\vec{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$

多次元尺度法（続き）

- 6 この距離行列に対して以下の行列 Q を両側からかける

$$Q = I - \frac{\vec{1} \cdot \vec{1}^T}{N}$$

このとき $\vec{1} \cdot \vec{1}^T = \begin{pmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{pmatrix}$ である。

これをヤング・ハウスホルダー変換という。 Q は N 個のすべての点を重心が原点にくるよう移動させる演算である。

多次元尺度法（続き）

- 7 この操作により
 $-\frac{1}{2}(QD^{(2)}Q) = QX^T XQ$ とかける
- 8 $D^{(2)}$ はわかっているので $QD^{(2)}Q$ を計算できる。
ここでの対角行列を逆行列とするような固有ベクトルを選ぶことができる（証明は略）
- 9 ただし今のように距離のみが分かっている場合は、 N 点のうち独立なものは $N - 1$ 個しかないので、固有値の 1 つは 0 になることに注意

多次元尺度法（続き）

10 よって以下のように変形できる

$$-\frac{1}{2}(QD^{(2)}Q) = P \begin{pmatrix} \lambda_1 & \dots & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & \lambda_{n-1} & 0 \\ 0 & \dots & 0 & 0 \end{pmatrix} P^T$$

$$11 \quad Z = P \begin{pmatrix} \sqrt{\lambda_1} & \dots & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & \sqrt{\lambda_{n-1}} & 0 \\ 0 & \dots & 0 & 0 \end{pmatrix}$$

とすると $QX(QX^T) = XX^T$ であり Z を解（の1つ）とできる

多次元尺度法（具体例）

- 1 4点 $x_1 = (3, 3)$, $x_2 = (3, 5)$, $x_3 = (5, 5)$, $x_4 = (5, 3)$ について、座標はわからず以下の距離行列のみが与えられていると考える。

$$D^{(2)} = \begin{pmatrix} 0 & 4 & 8 & 4 \\ 4 & 0 & 4 & 8 \\ 8 & 4 & 0 & 4 \\ 4 & 8 & 4 & 0 \end{pmatrix}$$

- 2 これにヤング・ハウスホルダー変換をすると

$$-\frac{1}{2}(QD^{(2)}Q) = \begin{pmatrix} 2 & 0 & -2 & 0 \\ 0 & 2 & 0 & -2 \\ -2 & 0 & 2 & 0 \\ 0 & -2 & 0 & 2 \end{pmatrix}$$

多次元尺度法（具体例）

- 3 この固有値は 4,4,0,0 となる。また、固有ベクトルは以下のようになる。

$$P = \begin{pmatrix} \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} & 0 \\ 0 & \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} & 0 \\ 0 & -\frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \end{pmatrix}$$

- 4 よって以下が成り立つ。

$$P^T(-\frac{1}{2}QD^{(2)}Q)P = \begin{pmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

多次元尺度法（具体例）

- 5 よって Z は QX に対応するものとして以下のよう
に求まる。

$$Z = \begin{pmatrix} \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} & 0 \\ 0 & \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} & 0 \\ 0 & -\frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$
$$= \begin{pmatrix} \sqrt{2} & 0 & 0 & 0 \\ 0 & \sqrt{2} & 0 & 0 \\ -\sqrt{2} & 0 & 0 & 0 \\ 0 & -\sqrt{2} & 0 & 0 \end{pmatrix}$$

多次元尺度法（具体例）

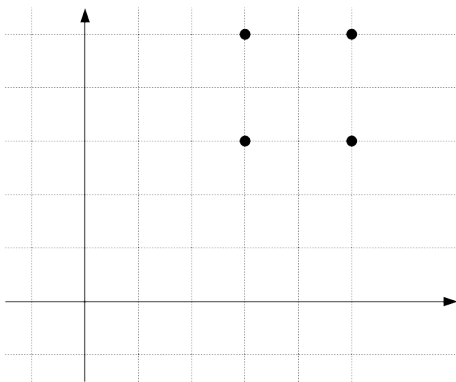
- 6 固有値が0に対応する部分を無視すると、
 $y_1(\sqrt{2}, 0), y_2(0, \sqrt{2}), y_3(-\sqrt{2}, 0), y_4(0, -\sqrt{2})$
の4点が求める座標となる
- 7 位置は移動し、さらに回転もしているが
ことに注意

多次元尺度法（具体例）

- 6 固有値が0に対応する部分を無視すると、
 $y_1(\sqrt{2}, 0), y_2(0, \sqrt{2}), y_3(-\sqrt{2}, 0), y_4(0, -\sqrt{2})$
の4点が求める座標となる
- 7 位置は移動し、さらに回転もしているが
距離は保存されている ことに注意

多次元尺度法（具体例）

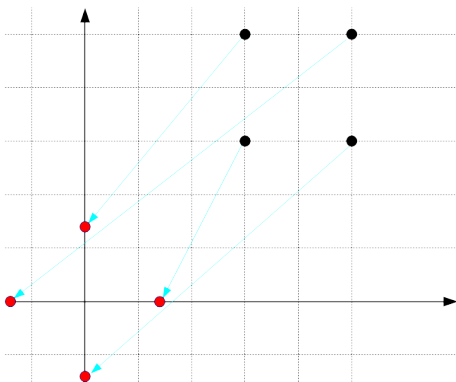
■ $x_1(3, 3), x_2(3, 5), x_3(5, 5), x_4(5, 3)$



多次元尺度法（具体例）

■ $x_1(3, 3), x_2(3, 5), x_3(5, 5), x_4(5, 3)$

■ $\Rightarrow y_1(\sqrt{2}, 0), y_2(0, \sqrt{2}), y_3(-\sqrt{2}, 0), y_4(0, -\sqrt{2})$



多次元尺度法（R実践）

- 大阪環状線の所要時間（内・外の短い方）を距離とする
- 外回り

	大阪	京橋	鶴橋	天王	新今	西九
大阪	0	7	16	23	27	38
京橋	38	0	9	16	20	31
鶴橋	29	36	0	7	11	22
天王寺	22	29	36	0	4	15
新今宮	18	25	34	41	0	11
西九条	7	14	23	30	34	0

多次元尺度法（R実践）

■ 内回り

	大阪	京橋	鶴橋	天王	新今	西九
大阪	0	38	29	22	18	7
京橋	7	0	36	29	25	14
鶴橋	16	9	0	38	34	23
天王寺	23	16	7	0	41	30
新今宮	27	20	11	4	0	34
西九条	38	31	22	15	11	0

多次元尺度法（R実践）

- 大阪環状線の所要時間（内・外の短い方）は以下になる

	大阪	京橋	鶴橋	天王	新今	西九
大阪	0	7	16	22	18	7
京橋	7	0	9	16	20	14
鶴橋	16	9	0	7	11	22
天王寺	22	16	7	0	4	15
新今宮	18	20	11	4	0	11
西九条	7	14	22	15	11	0

多次元尺度法（R実践）

- 1 以下のデータを “l:¥rdata¥mds01.csv” という名前で保存する。あるいは

<https://puffin.hannan-u.ac.jp/lect/dats/csvs/mds01.csv> からダウンロードする。

, "大阪", "京橋", "鶴橋", "天王", "新今", "西九"
"大阪", 0, 7, 16, 22, 18, 7
"京橋", 7, 0, 9, 16, 20, 14
"鶴橋", 16, 9, 0, 7, 11, 22
"天王寺", 22, 16, 7, 0, 4, 15
"新今宮", 18, 20, 11, 4, 0, 11
"西九条", 7, 14, 22, 15, 11, 0

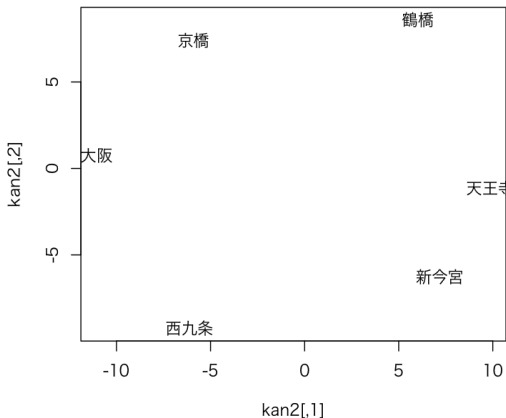
多次元尺度法（R実践）

2 R を立ち上げ、以下を実行する

```
kan <- read.csv('I:/rdata/mds01.csv',  
  ,header=T,row.names=1) # 本当は1行  
kan1 <- as.dist(kan) # 下の準備  
kan2 <- cmdscale(kan1) # MDS 処理  
plot(kan2,type="n")  
text(kan2,rownames(kan2))
```

多次元尺度法（R 実践）

☆ 出力結果



おしまい

質問があれば、お気軽に！