データの分析と知識発見

Introduction to Data Analysis

今回の構成

距離の公理について理解する

古典的多次元尺度法について学ぶ

Rを用いて古典的多次元尺度法を行う

距離の公理

非負性	$d(p, q) \geq 0$
非退化性	$p = q \iff d(p, q) = 0$
対称性	$d(p, q) = d(q, p)$
三角不等式	$d(p, q) \leq d(p, r) + d(r, q)$

古典的多次元尺度法

距離行列を

$$D^{(2)}=\begin{pmatrix} 0 & d_{12}^2 & \cdots & d_{1N}^2 \\ d_{21}^2 & 0 & \cdots & d_{2N}^2 \\ d_{31}^2 & d_{32}^2 & 0 & \cdots \\ \vdots & \vdots & \ddots & \\ d_{N1}^2 & d_{N2}^2 & \cdots & 0 \end{pmatrix}$$

とすると, $\boldsymbol{x}_i$, $\boldsymbol{x}_j$ を用いて

$$\begin{aligned} d_{ij}^2 &= \sum_{k=1}^{r}(x_{ik}-x_{jk})^2 \\ &= \boldsymbol{x}_i\cdot\boldsymbol{x}_i+\boldsymbol{x}_j\cdot\boldsymbol{x}_j-2\boldsymbol{x}_i\cdot\boldsymbol{x}_j \end{aligned}$$

行列で書くと

$$\begin{aligned} X &= (\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N)^T \\ &= \begin{pmatrix} x_{11}\cdots x_{1r} \\ \cdots \\ x_{N1}\cdots x_{Nr} \end{pmatrix} \end{aligned}$$

$$\begin{aligned} D^{(2)} &= \mathrm{diag}(XX^T)\cdot(1\cdot 1^T)-2XX^T \\ &\quad +(1\cdot 1^T)\cdot\mathrm{diag}(XX^T) \end{aligned}$$

となる

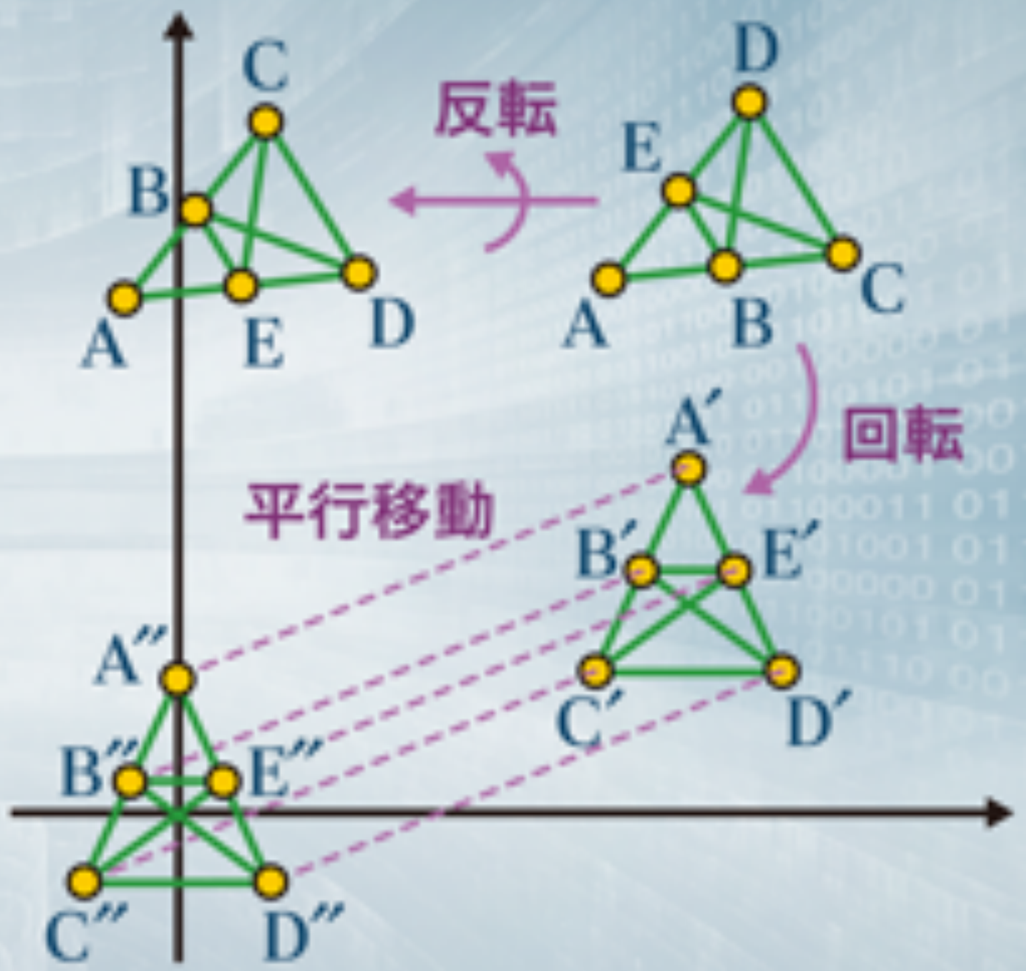
$$1\cdot 1^T=\begin{pmatrix} 1\cdots 1 \\ \ddots \\ 1\cdots 1 \end{pmatrix}$$

古典的多次元尺度法

回転、反転、平行移動しても
距離は変わらない

(答えは1つには決まらない)

$D^{(2)}$をもとに、直接Xを
求めるのではなく
$D^{(2)}$を満たすような
Xのうちの1つを求める



古典的多次元尺度法

ヤングハウスホルダー変換

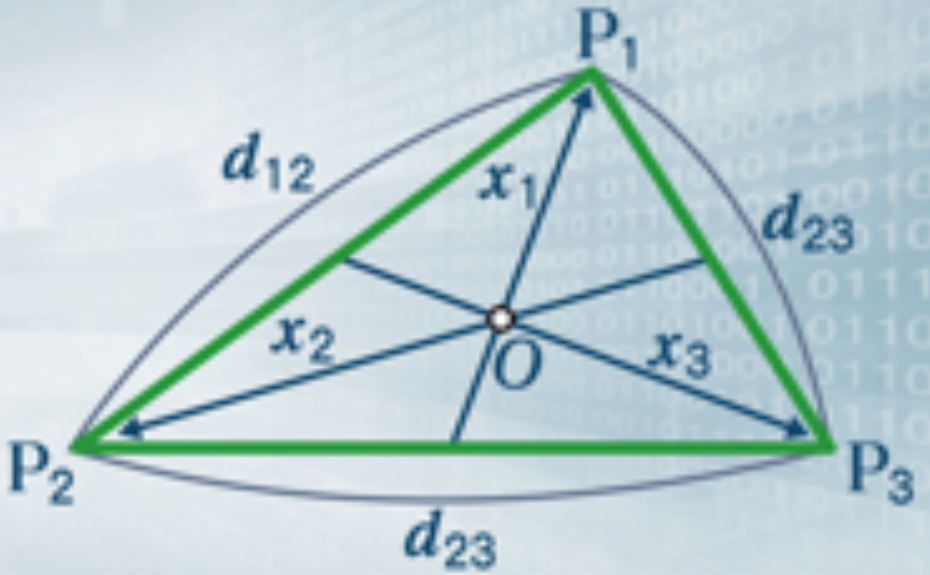
$$D^{(2)} = \mathrm{diag}(XX^T)\cdot(1\cdot 1^T) - 2XX^T + (1\cdot 1^T)\cdot \mathrm{diag}(XX^T)$$

という行列に対して

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

$$1\cdot 1^T = \begin{pmatrix} 1 & \cdots & 1 \\ & \ddots & \\ 1 & \cdots & 1 \end{pmatrix}$$

$$-\frac{1}{2}(QD^{(2)}Q) = QXX^TQ$$



平行移動した成分の内積を求めることができた

古典的多次元尺度法

$$-\frac{1}{2}(QD^{(2)}Q)=QXX^TQ$$

対角化すると、固有値がすべて0以上であれば

$$QXX^TQ=P\begin{pmatrix}\sqrt{\lambda_1} & \cdots & \cdots & 0\\ \vdots & \ddots & \vdots & \vdots\\ 0 & \cdots & \sqrt{\lambda_{N-1}} & 0\\ 0 & \cdots & 0 & 0\end{pmatrix}\begin{pmatrix}\sqrt{\lambda_1} & \cdots & \cdots & 0\\ \vdots & \ddots & \vdots & \vdots\\ 0 & \cdots & \sqrt{\lambda_{N-1}} & 0\\ 0 & \cdots & 0 & 0\end{pmatrix}P^T$$

中心が原点に来るように移動した座標の1つとして

$$Z=P\begin{pmatrix}\sqrt{\lambda_1} & \cdots & 0 & 0\\ \vdots & \ddots & \vdots & \vdots\\ 0 & \cdots & \sqrt{\lambda_{N-1}} & 0\\ 0 & \cdots & 0 & 0\end{pmatrix}$$

を選ぶことができる

多次元尺度法の計算（cmdscale）

```
> yamate0 <- read.csv ("ch101.csv", header=T, row.names=1)
> yamate1 <- as.dist (yamate0)
```

```
> yamate2 <- cmdscale (yamate1, eig=T, k=2)
```

```
> plot (yamate2$points, type="n")
> text (yamate2$points, rownames (yamate2$points))
```

多次元尺度法の計算（cmdscale）

所要時間データ

	1	2	3	・・・	12	13	14
1. 品川	0	7	12	・・・	14	10	7
2. 目黒	7	0	5	・・・	21	17	14
3. 渋谷	12	5	0	・・・	26	22	19
4. 原宿	14	7	2	・・・	28	24	21
・・・				・・・			
12. 秋葉原	14	21	26	・・・	0	4	7
13. 東京	10	17	22	・・・	4	0	3
14. 新橋	7	14	19	・・・	7	3	0

多次元尺度法の計算（cmdscale）

所要時間データ

CSVファイル

,品川,目黒,渋谷,原宿,新宿,高田馬場･･･
品川, 0 , 7 ,12,14,18,22,･･･
目黒, 7 , 0 , 5 , 7 ,11,15,･･･
渋谷,12, 5 , 0 , 2 , 6 ,10,･･･
原宿,14, 7 , 2 , 0 , 4 , 8 ,･･･
新宿,18,11, 6 , 4 , 0 , 4 ,･･･
高田馬場,22,15,10, 8 , 4 , 0 ,･･･
池袋,25,18,13,11, 7 , 3 ,･･･
⋮

多次元尺度法の計算（cmdscale）

所要時間データ

	1	2	3	・・・	12	13	14
1. 品川	0	7	12	・・・	14	10	7
2. 目黒	7	0	5	・・・	21	17	14
3. 渋谷	12	5	0	・・・	26	22	19
4. 原宿	14	7	2	・・・	28	24	21
・・・				・・・			
12. 秋葉原	14	21	26	・・・	0	4	7
13. 東京	10	17	22	・・・	4	0	3
14. 新橋	7	14	19	・・・	7	3	0

古典的多次元尺度法

距離行列を

$$D^{(2)}=\begin{pmatrix} 0 & d_{12}^2 & \cdots & d_{1N}^2 \\ d_{21}^2 & 0 & \cdots & d_{2N}^2 \\ d_{31}^2 & d_{32}^2 & 0 & \cdots \\ \vdots & \vdots & \ddots & \\ d_{N1}^2 & d_{N2}^2 & \cdots & 0 \end{pmatrix}$$

とすると, $\boldsymbol{x}_i$, $\boldsymbol{x}_j$ を用いて

$$\begin{aligned} d_{ij}^2 &= \sum_{k=1}^{r}(x_{ik}-x_{jk})^2 \\ &= \boldsymbol{x}_i\cdot\boldsymbol{x}_i+\boldsymbol{x}_j\cdot\boldsymbol{x}_j-2\boldsymbol{x}_i\cdot\boldsymbol{x}_j \end{aligned}$$

行列で書くと

$$\begin{aligned} X &= (\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N)^T \\ &= \begin{pmatrix} x_{11}\cdots x_{1r} \\ \cdots \\ x_{N1}\cdots x_{Nr} \end{pmatrix} \end{aligned}$$

$$\begin{aligned} D^{(2)} &= \mathrm{diag}(XX^T)\cdot(1\cdot 1^T)-2XX^T \\ &\quad +(1\cdot 1^T)\cdot\mathrm{diag}(XX^T) \end{aligned}$$

となる

$$1\cdot 1^T=\begin{pmatrix} 1\cdots 1 \\ \ddots \\ 1\cdots 1 \end{pmatrix}$$

古典的多次元尺度法

ヤングハウスホルダー変換

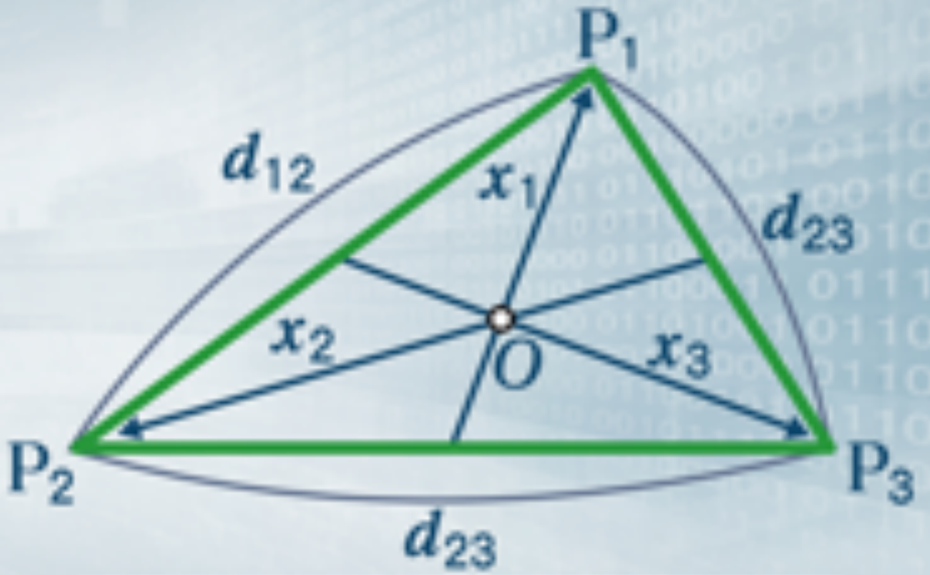
$$D^{(2)} = \mathrm{diag}(XX^T)\cdot(1\cdot 1^T) - 2XX^T + (1\cdot 1^T)\cdot \mathrm{diag}(XX^T)$$

という行列に対して

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

$$1\cdot 1^T = \begin{pmatrix} 1 & \cdots & 1 \\ & \ddots & \\ 1 & \cdots & 1 \end{pmatrix}$$

$$-\frac{1}{2}(QD^{(2)}Q) = QXX^TQ$$



平行移動した成分の内積を求めることができた

古典的多次元尺度法

距離行列を

$$D^{(2)}=\begin{pmatrix} 0 & d_{12}^2 & \cdots & d_{1N}^2 \\ d_{21}^2 & 0 & \cdots & d_{2N}^2 \\ d_{31}^2 & d_{32}^2 & 0 & \cdots \\ \vdots & \vdots & \ddots & \\ d_{N1}^2 & d_{N2}^2 & \cdots & 0 \end{pmatrix}$$

とすると, x_i, x_j を用いて

$$d_{ij}^2=\sum_{k=1}^{r}(x_{ik}-x_{jk})^2$$
$$=x_i\cdot x_i+x_j\cdot x_j-2x_i\cdot x_j$$

行列で書くと

$$X=(x_1,\cdots,x_N)^T$$
$$=\begin{pmatrix} x_{11}\cdots x_{1r} \\ \cdots \\ x_{N1}\cdots x_{Nr} \end{pmatrix}$$

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

となる

$$1\cdot 1^T=\begin{pmatrix} 1\cdots 1 \\ \ddots \\ 1\cdots 1 \end{pmatrix}$$

古典的多次元尺度法

ヤングハウスホルダー変換

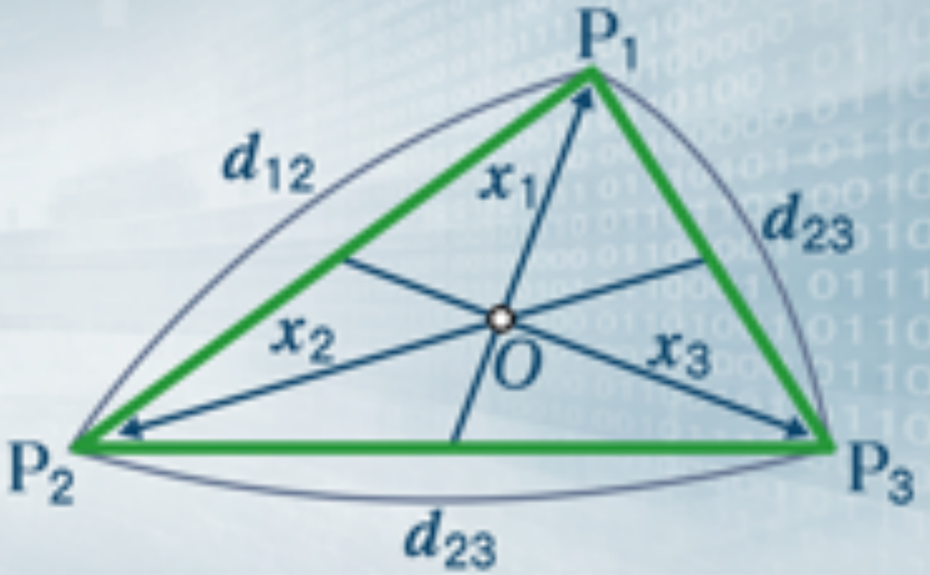
$$D^{(2)} = \mathrm{diag}(XX^T)\cdot(1\cdot 1^T) - 2XX^T + (1\cdot 1^T)\cdot \mathrm{diag}(XX^T)$$

という行列に対して

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

$$1\cdot 1^T = \begin{pmatrix} 1 & \cdots & 1 \\ & \ddots & \\ 1 & \cdots & 1 \end{pmatrix}$$

$$-\frac{1}{2}(QD^{(2)}Q) = QXX^TQ$$



平行移動した成分の内積を求めることができた