Mathematics, Pusan National University

Linear Algebra and Learning from Data

IV.10 Distance Matrices

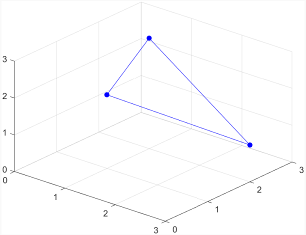
Taehyeong Kim
th_kim@pusan.ac.kr

December 30, 2020

Content

Suppose n points are at positions $\mathbf{x}_1, \ldots, \mathbf{x}_n$ in d-dimensional space. Distance matrix D is $n \times n$ matrix and $D_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|^2$ It is easy to see that D is symmetric.

Example 1.1



Let $\mathbf{x}_1 = (1,2,3), \mathbf{x}_2 = (3,2,1), \mathbf{x}_3 = (1,1,2)$, then the distance matrix D is

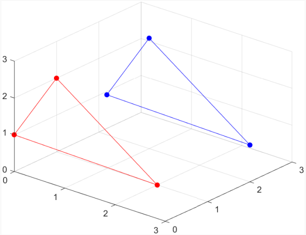
$$D = \begin{bmatrix} 0 & 8 & 2 \\ 8 & 0 & 6 \\ 2 & 6 & 0 \end{bmatrix}$$

Question

Can we recover the positions $\mathbf{x}_1, \ldots, \mathbf{x}_n$ from the Euclidean distance matrix D?

Answer : No.

Example 1.2



Let $\mathbf{x}_1 = (1, 2, 3), \mathbf{x}_2 = (3, 2, 1), \mathbf{x}_3 = (1, 1, 2)$ and $\tilde{\mathbf{x}}_1 = (0, 1, 2), \tilde{\mathbf{x}}_2 = (2, 1, 0), \tilde{\mathbf{x}}_3 = (0, 0, 1)$, then the distance matrix D and $\tilde{D}$ are same

$$\tilde{D} = D = \begin{bmatrix} 0 & 8 & 2 \\ 8 & 0 & 6 \\ 2 & 6 & 0 \end{bmatrix}$$

Another Question

Are there always positions $\mathbf{x}_1, \ldots, \mathbf{x}_n$ consistent with the distance matrix D?

Answer : Yes!

Application of distance matrix

1. Wireless sensor networks
2. Shapes of molecules
3. Machine learning

Positions X from Distances D

Let $X = \begin{bmatrix} | & & | \\ \mathbf{x}_1 & \cdots & \mathbf{x}_n \\ | & & | \end{bmatrix}$, then $X^T X = \begin{bmatrix} — & \mathbf{x}_1^T & — \\ & \vdots & \\ — & \mathbf{x}_n^T & — \end{bmatrix} \begin{bmatrix} | & & | \\ \mathbf{x}_1 & \cdots & \mathbf{x}_n \\ | & & | \end{bmatrix}$,

$$\|\mathbf{x}_i - \mathbf{x}_j\|^2 = (\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{x}_i - \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_i - \mathbf{x}_i^T \mathbf{x}_j - \mathbf{x}_j^T \mathbf{x}_i + \mathbf{x}_j^T \mathbf{x}_j \tag{1}$$

We can connect (1) and $X^T X$.

- Let $G = X^T X$, then the first term and last term $\mathbf{x}_k^T \mathbf{x}_k$ are in $\texttt{diag}(G)$.
- And middle terms of (1) $-\mathbf{x}_i^T \mathbf{x}_j - \mathbf{x}_j^T \mathbf{x}_i = -2\mathbf{x}_i^T \mathbf{x}_j$ is in $-2G$.

Then by using columns vector of n ones $\mathbf{1}$,

$$D = \mathbf{1}\texttt{diag}(G)^T - 2G + \texttt{diag}(G)\mathbf{1}^T \tag{2}$$

Positions X from Distances D

Example 2.1

Let $\mathbf{x}_1 = (1,2,3), \mathbf{x}_2 = (3,2,1), \mathbf{x}_3 = (1,1,2)$, then $X = \begin{bmatrix} 1 & 3 & 1 \\ 2 & 2 & 1 \\ 3 & 1 & 2 \end{bmatrix}$, $G = X^T X = \begin{bmatrix} 14 & 10 & 9 \\ 10 & 14 & 7 \\ 9 & 7 & 6 \end{bmatrix}$.

$$\mathbf{x}_1^T \mathbf{x}_1 = 14, \quad \mathbf{x}_2^T \mathbf{x}_2 = 14, \quad \mathbf{x}_3^T \mathbf{x}_3 = 6$$

$$\mathbf{x}_1^T \mathbf{x}_2 = 10, \quad \mathbf{x}_1^T \mathbf{x}_3 = 9, \quad \mathbf{x}_2^T \mathbf{x}_3 = 7$$

$$\mathbf{1}\mathtt{diag}(G)^T - 2G + \mathtt{diag}(G)\mathbf{1}^T = \begin{bmatrix} 0 & 8 & 2 \\ 8 & 0 & 6 \\ 2 & 6 & 0 \end{bmatrix} = D$$

Positions X from Distances D

Our problem is to recover G from D.

If G is positive semidefinite, we can find X with $X^T X = G$.
Since $\mathbf{1}\texttt{diag}(G)^T$ and $\texttt{diag}(G)\mathbf{1}^T$ terms in (2) are rank 1 matrices, D has rank $d+1$ or $d+2$.

Now we solve equation (2) for G.

1. Place first point at the origin: $\mathbf{x}_1 = \mathbf{0}$.
2. Then every $\|\mathbf{x}_i - \mathbf{x}_1\| = \|\mathbf{x}_i\|$.
3. Then the first column $\mathbf{d}_1$ of D must be exactly the same as $\texttt{diag}(X^T X) = \texttt{diag}(G) = (\|\mathbf{x}_1\|^2, \|\mathbf{x}_2\|^2, \ldots, \|\mathbf{x}_n\|^2)$.

$$\texttt{diag}(G) = \mathbf{d}_1 \quad \text{and} \quad G = -\frac{1}{2}(D - \mathbf{1}\mathbf{d}_1^T - \mathbf{d}_1\mathbf{1}^T). \tag{3}$$

Positions X from Distances D

Theorem 2.2

A symmetric matrix D with zero diagonal is a Euclidian distance matrix if and only is $\mathbf{x}^T D\mathbf{x} \leq 0$ for every vector with $\sum x_i = 0$ (so $\mathbf{1}^T\mathbf{x} = 0$).

We can solve $X^T X = G$ by several factorizations.

- Eigenvalue decomposition
  - If $G = Q\Lambda Q^T$, then $X = \sqrt{\Lambda}Q^T$.
- Cholesky factorization
  - If $G = U^T U$, then $X = U$.

Centering X or Rotating to Match Anchor Points

What is centering?

Centering : Often we might prefer to place the **centroid** of the $\mathbf{x}$'x at the origin. The centroid of $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is just the average of those vectors:

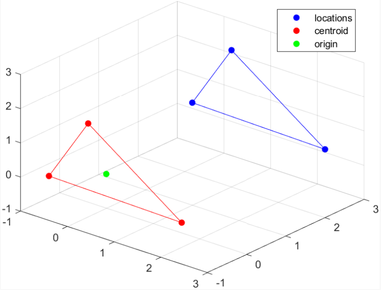
$$\textbf{Cdntroid} \quad c = \frac{1}{n}(\mathbf{x}_1 + \cdots + \mathbf{x}_n) = \frac{1}{n}X\mathbf{1}. \tag{4}$$

Just miltiply any position matrix X by the matrix $I - \frac{1}{n}\mathbf{1}\mathbf{1}^T$ to put the centroid at $\mathbf{0}$.

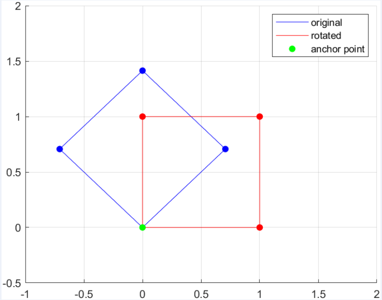
Centering X or Rotating to Match Anchor Points

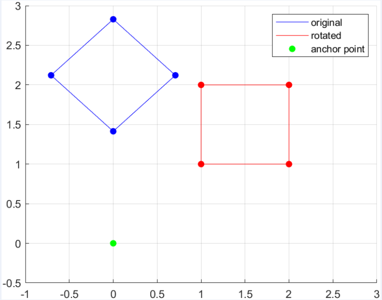
Example 3.1

Let $X = \begin{bmatrix} 1 & 3 & 1 \\ 2 & 2 & 1 \\ 3 & 1 & 2 \end{bmatrix}$, then $X(I - \frac{1}{n}\mathbf{1}\mathbf{1}^T) = \begin{bmatrix} -\frac{2}{3} & \frac{4}{3} & -\frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} & -\frac{2}{3} \\ 1 & -1 & 0 \end{bmatrix}$, average of columns is $\mathbf{0}$(origin).



What is anchor point?





Centering X or Rotating to Match Anchor Points

The best orthogonal matrix Q solves the Procrustes problem in Section IV.9.

Orthogonal Procrustes problem

$$R = \arg\min_{Q} \|X - YQ\|_F^2 \quad \text{subject to} \quad Q^T Q = I \tag{5}$$

(5) can be equivalent to find an orthogonal matrix R which most closely maps X to Y.

Problems

Problem 1

$\|\mathbf{x}_1 - \mathbf{x}_2\|^2$ and $\|\mathbf{x}_2 - \mathbf{x}_3\|^2$ and $\|\mathbf{x}_1 - \mathbf{x}_3\|^2$ will violate the triangle inequality. Construct G and comfirm that it is not positive semidefinite : no solution X to $G = X^T X$.

Problems

Problem 2

$||\mathbf{x}_1 - \mathbf{x}_2||^2 = 9$ and $||\mathbf{x}_2 - \mathbf{x}_3||^2 = 16$ and $||\mathbf{x}_1 - \mathbf{x}_3||^2 = 25$ do satisfy the triangle inequality $3 + 4 > 5$. Construct G and find points $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ that match these distances.

Problems

Problem 3

If all $\|\mathrm{x}_i - \mathrm{x}_j\|^2 = 1$ for $\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3, \mathrm{x}_4$, find G and then X. The points lie in $\mathbb{R}^d$ for which dimension d?

Problems

Thank you!