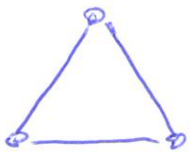
FBDA, 23.11.2018

4.2. Multidimensional Scaling

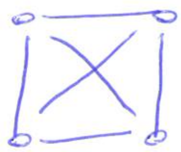
Problem 1) $\delta_{ij} = 1$ if $i \neq j$ $\delta_{ij} = 0$ if $i = j$ $i,j \in \{1,2,3\}$

$$\Delta_1 = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \qquad \delta_{ij}$$



$$\Delta_2 = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

$$\Delta_3 = \begin{pmatrix} 0 & 1 & \sqrt{2} & 1 \\ 1 & 0 & 1 & \sqrt{2} \\ \sqrt{2} & 1 & 0 & 1 \\ 1 & \sqrt{2} & 1 & 0 \end{pmatrix}$$



Given n Objects $O_1, O_2, \dots, O_n$ and pairwise dissimilarities δ_{ij} between objects i and j. We assume $\delta_{ij} = \delta_{ji}$ and $\delta_{ii} = 0$. and $\delta_{ij} \geq 0$. for all $i,j = 1, \dots, n$.

Define $\Delta = (\delta_{ij})_{1 \le i,j \le n}$ as the dissimilarity matrix

and $U_n = \left\{ \Delta = (\delta_{ij})_{1 \le i,j \le n} \;\middle|\; \begin{array}{l} \delta_{ij} = \delta_{ji} \geq 0 \\ \delta_{ii} = 0 \text{ for all } i,j \end{array} \right\}$

the set of all dissimilarity matrices.

Objective. Find n points $x_1, \dots, x_n$ in a Euclidean space, typically $\mathbb{R}^K$, such that the distances $\| x_i - x_j \|$ fit the dissimilarity δ_{ij}.

Example: Towns, δ_{ij} is the driving time from town i to town j. $\longrightarrow$ Find an embedding in $\mathbb{R}$.

Notation $\quad X = [x_1 \cdots x_n]^T \in \mathbb{R}^{n \times K}$

$$d_{ij}(X) = \| x_i - x_j \| \quad D(X) = (d_{ij}(X))_{1 \le i,j \le n}$$

$$\Delta^{(q)} = (\delta_{ij}^{q})_{ij} \text{ and } D^{(q)}(X) = (d_{ij}^{q}(X)).$$

Optimization problem: given q

$$\min_{X \in \mathbb{R}^{n \times K}} \| \Delta^{(q)} - D^{(q)}(X) \|$$

4.2.1. Characterizing Euclidean Distance Matrices

$\Delta = (\delta_{ij})_{ij} \in U_n$ is called Euclidean matrix

(or it has a Euclidean embedding in $\mathbb{R}^k$)

if there are $x_1, \dots, x_n \in \mathbb{R}^k$ such that

$$\delta_{ij} = \| x_i - x_j \| \qquad \forall i, j$$

where $\|y\| = \sqrt{\sum_{i=1}^{k} y_i^2}$.

$$E_n = I_n - \frac{1}{n} 1_n 1_n^T$$

Thm 4.3. $\Delta \in U_n$ has a Euclidean Embedding in $\mathbb{R}^K$ if and only if $-\frac{1}{2} E_n \Delta^{(2)} E_n$ is nonnegative definite and $rK(E_n \Delta^{(2)} E_n) \leq K$.

The least K which allows for an embedding is called dimensionality of Δ.

Proof. $X = (x_1 \cdots x_n)^T \in \mathbb{R}^{n \times K}$

$$\Delta = D(X) \qquad \Delta^{(q)} = D(X) \qquad q = 2.$$

$$\|x_i - x_j\|^2 = \|x_i\|^2 + \|x_j\|^2 - 2x_i^T x_j$$

$$x_i^T x_j = \frac{1}{2}\left(\|x_i\|^2 + \|x_j\|^2 - \|x_i - x_j\|^2\right)$$

$$-\frac{1}{2}\|x_i - x_j\|^2 = -\frac{1}{2}\|x_i\|^2 - \frac{1}{2}\|x_j\|^2 + x_i^T x_j$$

$\xrightarrow{\text{generalize}}$ $$-\frac{1}{2} D^{(2)}(X) = XX^T - 1\hat{x}^T - \hat{x}1^T$$

$$\hat{x} = (x_1^T x_1, x_2^T x_2, \dots, x_n^T x_n)^T$$

$$\Rightarrow \quad -\frac{1}{2} E_n D^{(2)}(X) E_n = E_n X X^T E_n \left.\begin{aligned} &- E_n 1 \hat{x}^T E_n \\ &- E_n \hat{x} 1^T E_n \end{aligned}\right\}$$

$(E_n 1 = 0)$

they are zero

$$\Rightarrow \quad -\frac{1}{2} E_n D^{(2)}(X) E_n = E_n X X^T E_n \quad (*)$$

$E_n = E_n^T$ $$= (\underline{E_n X})(\underline{E_n X})^T \succcurlyeq 0$$

n.n.d.

the reverse direction:

$$-\frac{1}{2} E_n \Delta^{(2)} E_n \succcurlyeq 0$$

$X \in \mathbb{R}^{n \times k}$

$$-\frac{1}{2} E_n \Delta^{(2)} E_n = V^T \Lambda V = XX^T$$

exercise

$$\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_n) \Rightarrow \Lambda = \Lambda^{1/2} \Lambda^{1/2}$$

$\lambda_i \geq 0$

$\mathrm{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots)$

$$rk\left(-\frac{1}{2}E_n\Delta^{(2)}E_n\right)=K \qquad \lambda_{K+1}, \lambda_{K+2}, \dots, \lambda_n=0$$

$$X\in\mathbb{R}^{n\times K} : \quad -\frac{1}{2}E_n\Delta^{(2)}E_n = XX^T$$

$$XX^T=-\frac{1}{2}E_n\Delta^{(2)}E_n$$

$$X^TE_n = X^T$$

$$X: \quad -\frac{1}{2}E_n\Delta^{(2)}E_n = V^T\Lambda V$$

$$\xrightarrow{\text{hint}} X = V^T\Lambda^{1/2} \qquad v_i \text{ eigenvectors}$$

$$\left(-\frac{1}{2}E_n\Delta^{(2)}E_n\right)v_i=\lambda_i v_i \Rightarrow v_i\in Im(E_n)$$

$$\Rightarrow E_n v_i = v_i \Rightarrow X^TE_n = X^T.$$

$$-\frac{1}{2}E_n D^{(2)}(X)E_n = E_nXX^TE_n = XX^T = -\frac{1}{2}E_n\Delta^{(2)}E_n.$$

$$\Rightarrow D^{(2)}(X)=\Delta$$

(exercise) $\rightarrow$ hollow matrices. $\square$

$-\frac{1}{2} E_n \Delta^{(2)} E_n$ in general is not non-negative definite. $\exists \lambda_i \quad \lambda_i < 0$

Thm. 4.4. Given $\Delta \in U_n$,

$$-\frac{1}{2} E_n \Delta^{(2)} E_n = V \operatorname{diag}(\lambda_1, \dots, \lambda_n) V^T$$ spectral decomp.

with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$,

$V = [v_1 \cdots v_n]$ corr. eigenvectors

$$\min_{X \in \mathbb{R}^{n \times k}} \| E_n (\Delta^{(2)} - D^{(2)}(X)) E_n \|$$

has a solution $X^* = (\sqrt{\lambda_1^+}\, v_1, \sqrt{\lambda_2^+}\, v_2, \dots, \sqrt{\lambda_k^+}\, v_k)$

$$\lambda^+ = \max\{\lambda, 0\}$$

Proof:

$$\min_{\substack{A \geq 0 \\ rk(A) \leq k}} \left\| -\frac{1}{2} E_n \Delta^{(2)} E_n - A \right\|^2 \quad (*)$$

reminder:

$$-\frac{1}{2} E_n D^{(2)}(X) E_n = E_n X X^T E_n \geq 0 \qquad rk\left(-\frac{1}{2} E_n D^{(2)}(X) E_n\right) \leq k$$

(Thm 2.6) : the min of $(*)$ is attained at

$$A^* = V \operatorname{diag}(\lambda_1^+, \lambda_2^+, \dots, \lambda_k^+, 0, \dots 0) V^T.$$

$$X^* = (\sqrt{\lambda_1^+}v_1, \sqrt{\lambda_2^+}v_2, \dots, \sqrt{\lambda_K^+}v_K) \in \mathbb{R}^{n \times K}$$

$$\Rightarrow A^* = X^* X^{*T}.$$

(exercise) $\square$

4.2.3. Non-linear dimensionality Reduction

→ Manifold learning

geodesics

ISOMAP (2000) – Science

Given data $x_1, \dots, x_n \in \mathbb{R}^p$ (lying on a manifold) e.g., the swiss roll.

→ first we construct a graph.

1) vertices v_i for each x_i

2) Connect v_i and v_j if $\|x_i - x_j\| < \varepsilon$
(another: connect to K-nearest neighbours)

3) For each pair, compute the shortest path between v_i, v_j: $d(v_i, v_j)$ (Dykstra's alg.)

4) Apply MDS on the basis of geodesic distances $\Delta = (d(v_i, v_j))_{1 \le i,j \le n}$.

Shortcomings of this approach:

- Choosing ε: difficult, not robust to noise perturbation

- Computational complexity: Dykstra's alg.

  MDS

- Very large distances may distort local neighborhoods