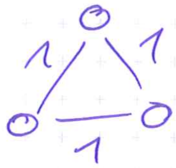
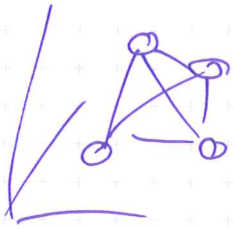
4.2 Multidimensional Scaling (MDS)



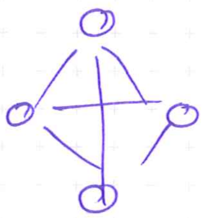
$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

Euclidean embedding in dim. 2 ?



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

Eucl. embedding in dim. 2 ? no!
in dim 3 ?



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

Eucl. emb. in dim. ?. ?

RWTH AACHEN UNIVERSITY

Given n objects $O_1, \dots, O_n$ and pairwise dissimilarities δ_{ij} between object i and j.

Assume that $\delta_{ij} = \delta_{ji} \geq 0$ and $\delta_{ii} = 0$, $1 \leq i,j \leq n$.

Define $\Delta = (\delta_{ij})_{1 \leq i,j \leq n}$ dissimilarity matrix.

and

$$\mathcal{U}_n = \{ \Delta = (\delta_{ij})_{1 \leq i,j \leq n} \mid \delta_{ij} = \delta_{ji} \geq 0, \delta_{ii} = 0 \ \forall i,j \}$$

the set of 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 dissimilarities δ_{ij} at its best.

Ex. Towns, δ_{ij} km between towns. Find a "map" in $\mathbb{R}^2$.

Notation: $X = (x_1, \dots, x_n)^T \in \mathbb{R}^{n \times k}$

$d_{ij}(X) = \|x_i - x_j\|$ distances

$D(X) = (d_{ij}(X))_{1 \leq i,j \leq n} \in \mathbb{R}^{n \times n}$

~~$\Delta^{(q)}$ and~~ $D^{(q)}(X) = (d_{ij}^q(X))_{1 \leq i,j \leq n}$

$\Delta^{(q)} = (\delta_{ij}^q)_{1 \leq i,j \leq n}$ (q th powers entrywise)

Thinking the Future
Zukunft denken

Optimization problem

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

Zero error case?

4.2.1 Characterizing Euclidean Distance Matrices

$\Delta = (\delta_{ij})$ is called Euclidean Dist. 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 $\delta_{ij}^2 = \|x_i - x_j\|^2 \; \forall i,j$.

where $\|\cdot\|$ denotes the Euclidean norm $\|y\| = \left(\sum_{i=1}^{k} y_i^2\right)^{1/2}$.

(i.e., $(*)$ has a solution with 0 error.)

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

$$= \begin{pmatrix} 1-\frac{1}{n} & -\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ \vdots & & & & \\ -\frac{1}{n} & -\frac{1}{n} & -\frac{1}{n} & \cdots & 1-\frac{1}{n} \end{pmatrix}$$

Th 4.3. $\Delta \in \mathcal{U}_n$ has a Euclidean embedding in $\mathbb{R}^k$ if and only if

$$-\frac{1}{2} E_n \Delta^{(2)} E_n \text{ is n.n.d.}$$

and $\operatorname{rk}(E_n \Delta^{(2)} E_n) \leq k$. The least k which allows for an embedding is called dimensionality of Δ. ⌟

Proof. Given $X \in \mathbb{R}^{n \times k}$, $X = (x_1, \dots, x_n)^T$. It holds

Ex $\quad -\frac{1}{2} D^{(2)}(X) = XX^T - 1\hat{x}^T - \hat{x}^T 1$, $\quad \hat{x} = \frac{1}{2}(x_1^T x_1, \dots, x_n^T x_n)$

$$-\frac{1}{2} E_n D^{(2)}(X) E_n = E_n X X^T E_n \geq 0 \text{ and } \operatorname{rk}(E_n X X^T E_n) \leq k.$$

"$\Rightarrow$" Ass. $\exists$ Eucl. emb., i.e., $\Delta^{(2)} = D^{(2)}(X)$.

Then $\cancel{-\frac{1}{2}\Delta^{(2)}}$ $-\frac{1}{2} E_n \Delta^{(2)} E_n = -\frac{1}{2} E_n D^{(2)}(X) E_n \geq 0$, $\operatorname{rk}(\dots) \leq k$.

"$\Leftarrow$" Let $-\frac{1}{2} E_n \Delta^{(2)} E_n \geq 0$, $\operatorname{rk}(E_n \Delta^{(2)} E_n) \leq k$. Then there exists some $X \in \mathbb{R}^{n \times k}$ such that

Ex $$-\frac{1}{2} E_n \Delta^{(2)} E_n = XX^T \text{ and } X^T E_n = X^T.$$

$X = (x_1, \dots, x_n)^T$ is an appr. configuration, i.e.,

$$-\frac{1}{2} E_n \underline{\underline{D^{(2)}(X)}} E_n = E_n X X^T E_n = X X^T = -\frac{1}{2} E_n \underline{\underline{\Delta^{(2)}}} E_n$$

Ex It follows: $D^{(2)}(X) = \Delta^{(2)}$. ∎

4.2.2. The best Euclidean fit to a given dissimilarity matrix

$\|\cdot\|$ Frobenius norm, $\|A\| = \left(\sum_{i,j} a_{ij}^2\right)^{1/2}$

$\lambda^+ = \max\{\lambda, 0\}$ denotes the positive part.

Th. 4.4. Given $\Delta \in \mathcal{U}_n$.

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

with $\lambda_1 \geq \dots \geq \lambda_n$, $V = (v_1, \dots, v_n)$ orth. eigenvectors.

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

has a solution

$$X^* = \left(\sqrt{\lambda_1^+}\, v_1, \dots, \sqrt{\lambda_k^+}\, v_k\right) \in \mathbb{R}^{n \times k} \,\perp$$

Proof. $$\min_{A \geq, \operatorname{rk}(A) \leq k} \| -\frac{1}{2} E_n \Delta^{(2)} E_n - A \|^2 \qquad \text{(Th. 2.6)}$$

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

It holds

$$-\frac{1}{2} E_n D^{(2)}(X^*) E_n = E_n X^* X^{*T} E_n \qquad \text{(see above)}$$

$$= E_n (v_1, \dots, v_k) \operatorname{diag}(\lambda_1^+, \dots, \lambda_k^+) (v_1, \dots, v_k)^T E_n$$

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

so that the min is attained in the set

$$\left\{ -\frac{1}{2} E_n D^{(2)}(X) E_n \mid X \in \mathbb{R}^{n \times k} \right\}.$$

$\boxtimes$

4.2.3. Non-linear dimensionality reduction

ISOMAP (Tenebaum, deSilva, Langford, Science 290 (2000))

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

Generate a graph with vertices $v_i = x_i$
and link vertices v_i and v_j only if $\|x_i - x_j\| < \varepsilon$ (small)

Algorithm:

a) For each pair (v_i, v_j) compute the shortest path (Dykstra's algorithm)

The geodesic distance $d(v_i, v_j)$ can be taken as

- number of hops / links from v_i to v_j
- sum of $\|x_i - x_j\|$ on a shortest path

b) Apply MDS on the basis of geodesic distances

$$\Delta = \left(\delta(v_i, v_j)\right)_{1 \le i,j \le n}$$

Shortcomings:

- Very large distances may distort local neighborhoods
- Computational complexity: Dykstra, MDS.
- not robust to noise perturbation