MULTIDIMENSIONAL SCALING

MDS is a family of algorithms that attempt to

(A) solve the ________ ________ problem:

Given an $n \times n$ dissimilarity matrix $D = [d_{ij}]$, find a dimension p and points $x_1, \ldots, x_n \in \mathbb{R}^p$ such that

$$d_{ij} = \| x_i - x_j \|$$

(B) Applications!

-
-
-

Recall | A dissimilarity matrix should satisfy

- $d_{ij} \geq 0$
- $d_{ij} = d_{ji}$
- $d_{ii} = 0$

We do not require the triangle inequality.

In some cases, a Euclidean embedding may not exist. In other cases, the value of p may be too large for the intended application. Therefore ____________ solutions are also of interest.

Nomenclature |

- ____________ methods attempt to preserve all interpoint distances
- ____________ methods only attempt to preserve rank ordering

$\boxed{\text{Euclidean Distance Matrices}}$

Definition | An $n \times n$ matrix D is a Euclidean distance matrix if there exists p and $x_1, \ldots, x_n \in \mathbb{R}^p$ such that $d_{ij} = \| x_i - x_j \|$ for all i, j.

Theorem | Let D be an $n \times n$ dissimilarity matrix. Set

$$B = H \cdot A \cdot H$$

where

$$A = [a_{ij}], \qquad a_{ij} = -\tfrac{1}{2} d_{ij}^2$$

$$H = I - \tfrac{1}{n} \mathbb{1} \mathbb{1}^T$$

Then D is a Euclidean distance matrix

Ⓔ iff B is __________ ______ __________.

(continued)

Furthermore, suppose B is PSD with positive eigenvalues

$$\lambda_1 > \lambda_2 > \cdots > \lambda_p$$

and corresponding eigenvectors

$$u_1 = \begin{bmatrix} u_{11} \\ u_{12} \\ \vdots \\ \\ u_{1n} \end{bmatrix}, \quad \cdots, \quad u_p = \begin{bmatrix} u_{p1} \\ u_{p2} \\ \vdots \\ \\ u_{pn} \end{bmatrix}$$

normalized such that

$$u_k^T u_k = \lambda_k$$

Then the vectors

$$x_i :=$$

E

satisfy

In addition

$$\bar{x} =$$

<u>Proof</u> Suppose D is a EDM and let $x_1, \dots, x_n \in \mathbb{R}^p$ such that $d_{ij} = \|x_i - x_j\|$. It can be shown through straightforward algebra that

$$B = [b_{ij}]$$

where

$$b_{ij} = a_{ij} - \frac{1}{n}\sum_{k=1}^{n} a_{ik} - \frac{1}{n}\sum_{k=1}^{n} a_{kj} + \frac{1}{n^2}\sum_{k,l} a_{kl}$$

Substituting

$$a_{ij} = \frac{-1}{2}(x_i - x_j)^T (x_i - x_j)$$

gives

$$b_{ij} = (x_i - \bar{x})^T (x_j - \bar{x})$$

To see that B is PSD, let $f \in \mathbb{R}^n$ be arbitrary. Then

$$\begin{aligned} f^T B f &= \sum_{i=1}^{n} \sum_{j=1}^{n} f_i f_j b_{ij} \\ &= \sum_i \sum_j f_i f_j \langle x_i - \bar{x}, x_j - \bar{x} \rangle \\ &= \sum_i f_i \langle x_i - \bar{x}, \sum_j f_j (x_j - \bar{x}) \rangle \end{aligned}$$

$$= \left\langle \sum_i f_i (x_i - \bar{x}), \sum_j f_j (x_j - \bar{x}) \right\rangle$$

$$= \left\| \sum_i f_i (x_i - \bar{x}) \right\|^2 \geq 0$$

Now suppose B is PSD and let $x_1, \dots, x_n \in \mathbb{R}^p$ be as constructed. Let the eigenvalue decomposition of be

$$B = V \cdot \Lambda \cdot V^T$$

$$= U U^T$$

where

$$U = \begin{bmatrix} \sqrt{\lambda_1} v_1 & \cdots & \sqrt{\lambda_p} v_p & 0 & \cdots & 0 \end{bmatrix}$$

This shows that

$$b_{ij} = \langle x_i, x_j \rangle = x_i^T x_j$$

Finally, observe that

$$\begin{aligned} B \cdot \underline{1} &= HAH\underline{1} \\ &= HA\left(I - \tfrac{1}{n}\underline{1}\,\underline{1}^T\right)\underline{1} \\ &= H \cdot A \cdot 0 = 0 \end{aligned}$$

Therefore $\underline{1}$ is an eigenvector of B with eigenvalue 0. Hence $\underline{1}$ is orthogonal to $u_1, \dots, u_p$. That is,

$$u_k^T \underline{1} = \sum_{i=1}^{p} x_{ik} = 0$$

$$\Longrightarrow \bar{x} = [0 \;\; 0 \;\; \cdots \;\; 0]$$

$\blacksquare$

Therefore

$$(x_i - x_j)^T (x_i - x_j)$$

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

$$= b_{ii} - 2 b_{ij} + b_{jj}$$

$$= \left[a_{ii} - \frac{1}{n} \sum_k a_{ik} - \frac{1}{n} \sum_k a_{ki} + \frac{1}{n^2} \sum_{k,l} a_{kl} \right]$$

$$-2 \left[a_{ij} - \frac{1}{n} \sum_k a_{ik} - \frac{1}{n} \sum_k a_{kj} + \frac{1}{n^2} \sum_{k,l} a_{kl} \right]$$

$$+ \left[a_{jj} - \frac{1}{n} \sum_k a_{jk} - \frac{1}{n} \sum_k a_{kj} + \frac{1}{n^2} \sum_{k,l} a_{kl} \right]$$

$$= a_{ii} - 2 a_{ij} + a_{jj}$$

$$= 0 + d_{ij}^2 + 0$$

$$= d_{ij}^2$$

Classical MDS

Even if a dissimilarity matrix D cannot be embedded into p dimensions, the previous result suggests an approximate algorithm

CLASSICAL MDS

Input: D, desired dimension p

1. Form $B = HAH$

   where $A = (a_{ij})$, $a_{ij} = -d_{ij}^2$

2. Compute eigenvalue decomposition

$$B = V \Lambda V^T$$

   where $V = [v_1 \cdots v_n]$, $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$,

   $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$.

3. Set $u_i = \sqrt{\lambda_i}\, v_i$, $U = [u_1 \cdots u_p]$

Return: $x_i =$ ith row of U, $i = 1, \ldots, n$

In Matlab: cmdscale

(G) **Note** The algorithm cannot be applied if

Relation to PCA

Theorem Let $z_1, \dots, z_n \in \mathbb{R}^q$ and $D = (d_{ij})$ where $d_{ij} = \|z_i - z_j\|$. Let $x_1, \dots, x_n \in \mathbb{R}^p$ be the classical MDS embedding. Then

x_i = projection of $z_i - \bar{z}$ onto first p principal eigenvectors

Ref Mardia, Kent, & Bibby,
Multivariate Analysis, 1979

Stress Criteria

Another common approach to MDS is to

(H) minimize the ________ function

$$\sum_{i,j}$$

where w_{ij} are fixed. For example,

$$w_{ij} =$$

$$w_{ij} =$$

Stress criteria are typically minimized by gradient descent. The most common algorithm is called the majorization algorithm. For details, see Jan de Leeuw, "Convergence of the Majorization Method for Multidimensional Scaling," J. Classification 5:163-180 (1988)

In Matlab: mdscale

$\boxed{\text{Key}}$ A. Euclidean embedding B. extend algorithms to non-Euclidean data, visualization ($p = 1, 2, 3$), dim. reduction

C. approximate D. metric, nonmetric E. positive semi-definite

F. $x_i = (u_{1i} \ u_{2i} \cdots u_{pi})$, $d_{ij} = \|x_i - x_j\|$, $\bar{x} = 0$

G. $\lambda_k < 0$, for some k, $1 \le k \le p$.

H. Stress

$$\sum_{i,j} w_{ij} \left(d_{ij} - \|x_i - x_j\| \right)^2$$

$$w_{ij} = 1, \text{ or } d_{ij}^{-\alpha}, \ \alpha > 0$$