MULTIDIMENSIONAL SCALING

MDS is a family of algorithms that attempt to 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

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

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
**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

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

$$H = I - \frac{1}{n} 1_1^T$$

Then $D$ is a Euclidean distance matrix iff $B$ is

(continued)
Furthermore, suppose \( B \) is PSD with positive eigenvalues

\[
\lambda_1 > \lambda_2 > \ldots > \lambda_p
\]

and corresponding eigenvectors

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

normalized such that

\[
u_k^T u_k = \lambda_k
\]

Then the vectors

\[
\chi_i := B
\]

satisfy

In addition

\[
\overline{\chi} = \chi
\]
Proof. Suppose \( D \) is a EDM and let \( \chi_1, \ldots, \chi_n \in \mathbb{R}^p \) such that \( d_{ij} = \| \chi_i - \chi_j \| \). 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

\[
a_{ij} = -\frac{1}{2} (\chi_i - \chi_j)^T (\chi_i - \chi_j)
\]

gives

\[
b_{ij} = (\chi_i - \bar{x})^T (\chi_j - \bar{x})
\]

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

\[
f^T B f = \sum_{i=1}^{n} \sum_{j=1}^{n} f_i f_j b_{ij}
\]

\[
= \sum_{i} \sum_{j} f_i f_j \langle \chi_i - \bar{x}, \chi_j - \bar{x} \rangle
\]

\[
= \sum_{i} f_i \langle \chi_i - \bar{x}, \sum_{j} f_j (\chi_j - \bar{x}) \rangle
\]
\[
\begin{align*}
&= \left< \sum_i f_i(x_i - \bar{x}), \sum_j f_j(x_j - \bar{x}) \right> \\
&= \left\| \sum_i f_i(x_i - \bar{x}) \right\|^2 > 0
\end{align*}
\]

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

\[
B = V \cdot \Lambda \cdot V^T
\]

\[
\Lambda = UU^T
\]

where

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

This shows that

\[
b_{ij} = \left< x_i, x_j \right> = x_i^T x_j
\]
Finally, observe that

\[ B \cdot 1 = H A H \cdot 1 \]

\[ = H A (I - \frac{1}{n} 1 1^T) 1 \]

\[ = H \cdot A \cdot 0 = 0 \]

Therefore \( 1 \) is an eigenvector of \( B \) with eigenvalue 0. Hence \( 1 \) is orthogonal to \( u_1, \ldots, u_p \). That is,

\[ u_k^T \cdot 1 = \sum_{i=1}^{P} x_{ik} = 0 \]

\[ \Rightarrow \bar{x} = [0 \ 0 \ \ldots \ 0] \]
Therefore

\[(x_i - x_j) (x_i - x_j)\]

\[= x_i^T x_i - 2 x_i^T x_j + x_j^T x_j\]

\[= b_{ii} - 2 b_{ij} + b_{jj}\]

\[= \left[ a_{ii} - \frac{1}{h} \sum_k a_{ik} - \frac{1}{h} \sum_k a_{ki} + \frac{1}{h^2} \sum_{k,l} a_{kl} \right] - 2 \left[ a_{ij} - \frac{1}{h} \sum_k a_{ik} - \frac{1}{h} \sum_k a_{kj} + \frac{1}{h^2} \sum_{k,l} a_{kl} \right] + \left[ a_{jj} - \frac{1}{h} \sum_k a_{jk} - \frac{1}{h} \sum_k a_{kj} + \frac{1}{h^2} \sum_{k,l} a_{kl} \right]\]

\[= a_{ii} - 2 a_{ij} + a_{jj}\]

\[= 0 + d_{ij}^2 + b\]

\[= d_{ij}^2\]
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 $V = [v_1 \ldots v_n]$, $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$,
   $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$.

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

Return: $X_i = $ $i$th row of $U$, $i = 1, \ldots, n$

In Matlab: `cmdscale`
Note: The algorithm cannot be applied if

Relation to PCA

**Theorem** Let \( z_1, \ldots, z_n \in \mathbb{R}^q \) and \( D = (d_{ij}) \)
where \( d_{ij} = ||z_i - z_j||_1 \). Let \( x_1, \ldots, x_n \in \mathbb{R}^p \)
be the classical MDS embedding. Then \( x_i = \text{projection of } z_i - \bar{z} \text{ onto first } p \)
principal eigenvectors.

Ref: Mardia, Kent, & Bibby,

*Multivariate Analysis*, 1979
Stress Criteria

Another common approach to MDS is to minimize the function

\[
\sum_{i,j}
\]

where \(w_{ij}\) are fixed. For example,

\[
\begin{align*}
    w_{ij} &= \\
    w_{ij} &=
\end{align*}
\]

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
Key

A. Euclidean embedding B. extend algorithms to non-Euclidean data, visualization (p=1,2,3), dim. reduction
C. approximate D. metric, non-metric E. positive semi-definite

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

G. \( \lambda_k < 0 \), for some \( k, 1 \leq k \leq p \).

H. stress

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

\( w_{ij} = 1 \), or \( d_{ij}^{-\alpha}, \alpha > 0 \)