1 MDS

- Multidimensional scaling (MDS) is essentially a data reduction technique, with the aim to find a set of points in low dimension that approximate the possibly high-dimensional configuration represented by the original proximity matrix.
- Distance Matrix $D = (d_{ij})_{n\times n}$,

$$D = D^T,\ d_{ij} \geq 0,\ d_{ii} = 0,\ i, j = 1, 2, \ldots, n.$$

- Euclidean Distance Matrix $D = (d_{ij})_{n\times n}$,

$$d_{ij}^2 = (x_i - x_j)^T (x_i - x_j),\ i, j = 1, 2, \ldots, n. \tag{1}$$

2 CMDS: Classical multi-dimensional scaling

- From distance matrix D to derive B.
- To recover the coordinates X from B.
- Dissimilarities as Euclidean distances.
- How many dimensions?
- A practical algorithm for classical MDS.

2.1 From D to derive B

- First assume that $X \in \mathbb{R}^{n\times p}$ is known and $x_i = (x_{i1}, \ldots, x_{ip})^T$ (column vectors).
- We define $B = (b_{ij})_{n\times n}$ as the inner product matrix

$$b_{ij} = x_i^T x_j. \tag{2}$$

- Note that we assume that $\sum_{i=1}^{n} x_{ik} = 0$, for $k = 1, \ldots, p$.

- According to Equations (1) and (2), we have

$$
\begin{aligned}
&d_{ij}^2 = x_i^T x_i + x_j^T x_j - 2x_i^T x_j. \\
&\frac{1}{n}\sum_{i=1}^{n} d_{ij}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^T x_i + x_j^T x_j - 2\frac{1}{n}\sum_{i=1}^{n} x_i^T x_j = \frac{1}{n}\sum_{i=1}^{n} x_i^T x_i + x_j^T x_j. \\
&x_j^T x_j = \frac{1}{n}\sum_{i=1}^{n} d_{ij}^2 - \frac{1}{n}\sum_{i=1}^{n} x_i^T x_i. \\
&\frac{1}{n}\sum_{j=1}^{n} d_{ij}^2 = x_i^T x_i + \frac{1}{n}\sum_{j=1}^{n} x_j^T x_j = x_i^T x_i + \frac{1}{n}\sum_{i=1}^{n} x_i^T x_i. \\
&x_i^T x_i = \frac{1}{n}\sum_{j=1}^{n} d_{ij}^2 - \frac{1}{n}\sum_{i=1}^{n} x_i^T x_i. \\
&\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i^T x_i + \frac{1}{n}\sum_{j=1}^{n} x_j^T x_j\right) = \frac{2}{n}\sum_{i=1}^{n} x_i^T x_i. \\
&b_{ij} = x_i^T x_j = -\frac{1}{2}\left(d_{ij}^2 - x_i^T x_i - x_j^T x_j\right) \\
&= -\frac{1}{2}\left(d_{ij}^2 - \frac{1}{n}\sum_{j=1}^{n} d_{ij}^2 - \frac{1}{n}\sum_{i=1}^{n} d_{ij}^2 + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}^2\right) \\
&= -\frac{1}{2}\left(d_{ij}^2 - d_{i.}^2 - d_{.j}^2 + d_{..}^2\right).
\end{aligned}
$$

- Define matrix A as $A = (a_{ij})_{n\times n}$ with $a_{ij} = -\frac{1}{2}d_{ij}^2$, then we have $b_{ij} = a_{ij} - a_{i.} - a_{.j} + a_{..}$.
- The inner product matrix B is

$$B = HAH$$

  where H is the centering matrix,

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

  with $\mathbf{1} = (1, \ldots, 1)^T$.

  –
$$
\begin{aligned}
B &= HAH \\
&= (I - \frac{1}{n}\mathbf{1}\mathbf{1}^T)A(I - \frac{1}{n}\mathbf{1}\mathbf{1}^T) \\
&= (I - \frac{1}{n}\mathbf{1}\mathbf{1}^T)(A - \frac{1}{n}A\mathbf{1}\mathbf{1}^T) \\
&= A - \frac{1}{n}A\mathbf{1}\mathbf{1}^T - \frac{1}{n}\mathbf{1}\mathbf{1}^T A + \frac{1}{n^2}\mathbf{1}\mathbf{1}^T A\mathbf{1}\mathbf{1}^T.
\end{aligned}
$$

  –
$$A = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{n1} & \ldots & a_{nn} \end{pmatrix}, \mathbf{1}\mathbf{1}^T = \begin{pmatrix} 1 & \ldots & 1 \\ \vdots & \vdots & \vdots \\ 1 & \ldots & 1 \end{pmatrix}.$$

- $(A\mathbf{1}\mathbf{1}^T)_{ij} = \sum_{k=1}^{n} a_{ik} = n \times a_{i\cdot}$.
-
$$\mathbf{1}\mathbf{1}^T = \begin{pmatrix} 1 & \dots & 1 \\ \vdots & \vdots & \vdots \\ 1 & \dots & 1 \end{pmatrix}, A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix}.$$
- $(\mathbf{1}\mathbf{1}^T A)_{ij} = \sum_{k=1}^{n} a_{kj} = n \times a_{\cdot j}$.

2.2 To recover the coordinates X from B

Lemma 1. *Let $A \in \mathbb{R}^{m\times k}$ and $B \in \mathbb{R}^{k\times n}$. Suppose that B has full row rank, then*

$$\begin{aligned} &Range(AB) = Range(A). \quad \textit{Range space is a subspace.} \\ &rank(AB) = dim\left(Range(AB)\right) = dim\left(Range(A)\right) = rank(A). \end{aligned}$$

Proof. B has full row rank, that is $\text{rank}(B) = k$ and $\text{Range}(B) = \mathbb{R}^k$.

$$\begin{aligned} \text{Range}(AB) &= \{y \mid y = ABx, x \in \mathbb{R}^n\} \\ &= \{y \mid y = Az, z = Bx, x \in \mathbb{R}^n\} \\ &= \left\{y \mid y = Az, z \in \text{Range}(B) = \mathbb{R}^k\right\} \\ &= \left\{y \mid y = Az, z \in \mathbb{R}^k\right\} \\ &= \text{Range}(A). \end{aligned}$$

□

- $X \in \mathbb{R}^{n\times p}$ and $x_i = (x_{i1}, \dots, x_{ip})^T$.
- $\text{rank}(X^T) = \text{rank}(X) = p$, X^T is a full row rank matrix.
- The inner product matrix $B = (b_{ij})_{n\times n}$ where $b_{ij} = x_i^T x_j$.
- B can be expressed as
$$B = XX^T.$$
- $\text{rank}(B) = \text{rank}(XX^T) = \text{rank}(X) = p$.
- B is symmetric, positive semi-definite and of rank p, hence B has p non-negative eigenvalues and $n - p$ zero eigenvalues.
- B can be written in terms of its spectral decomposition,
$$B = V\Lambda V^T.$$
- $\Lambda = diag\left(\lambda_1, \dots, \lambda_n\right)$, the diagonal matrix of eigenvalues of B. $\lambda_1 \geq \dots \geq \lambda_n \geq 0$.
- $V = [v_1, \dots, v_n]$ is the matrix of corresponding eigenvectors, normalized such that $v_i^T v_i = 1$.

- Because of the $n-p$ zero eigenvalues, B can now be rewritten as

$$B = V_1 \Lambda_1 V_1^T.$$

$$\Lambda_1 = diag\left(\lambda_1, \ldots, \lambda_p\right), \quad V_1 = [v_1, \ldots, v_p].$$

- The coordinate matrix X is given by

$$X = V_1 \Lambda_1^{\frac{1}{2}},$$

  where $\Lambda_1^{\frac{1}{2}} = diag\left(\lambda_1^{\frac{1}{2}}, \ldots, \lambda_p^{\frac{1}{2}}\right)$.

- Thus the coordinates of the points have been recovered from the distances between the points.

2.3 Dissimilarities as Euclidean distances

Suppose we have some dissimilarities $\{\delta_{ij}\}$ instead of distances $\{d_{ij}\}$, then we doubly produce matrix B as just described. Can this B give rise to a configuration of points in Euclidean space?

- Case with B is positive semi-definite of rank p.
  - A configuration in p dimensional Euclidean space can be found.
- Case with B is not positive semi-definite.
  - a constant can be added to all the dissimilarities (except the self-dissimilarities $\{\delta_{ii}\}$) which will then make B PSD.

$$\delta'_{ij} = \delta_{ij} + c(1 - \delta^{ij}),$$

    where c is an appropriate constant and $\delta^{ij} = 1$ if $i = j$ and zero otherwise.
  - Additive constant problem.

2.3.1 Case with B is positive semi-definite

-

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

  where $\Lambda = diag\left(\lambda_1, \ldots, \lambda_p\right)$ and $X = [x_i]^T$.

- Now the distance between the ith and jth points of configuration is given by $(x_i - x_j)^T(x_i - x_j)$, and hence

$$\begin{aligned}
(x_i - x_j)^T(x_i - x_j) &= x_i^T x_i + x_j^T x_j - 2x_i^T x_j \\
&= b_{ii} + b_{jj} - 2b_{ij} \\
&= (a_{ii} - a_{i.} - a_{.i} + a_{..}) + (a_{jj} - a_{j.} - a_{.j} + a_{..}) - 2(a_{ij} - a_{i.} - a_{.j} + a_{..}) \\
&= -2a_{ij} = \delta_{ij}^2.
\end{aligned}$$

- Note that $a_{ij} = -\frac{1}{2}d_{ij}^2$, $a_{ii} = 0$ and $a_{i.} = a_{.i}$.
- Thus the distance between the ith and jth points in the Euclidean space is equal to the original dissimilarity δ_{ij}.
- D is the Euclidean distance matrix equivalent to B is PSD.

2.4 How many dimensions?

- Multidimensional scaling (MDS) is essentially a data reduction technique.
- If we would like to use $m < p$ dimensions such that could make a good representation.
- Simply choosing the first m eigenvalues and eigenvectors of B will give a small dimensional space for the points.

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i^T x_i + \frac{1}{n}\sum_{j=1}^{n} x_j^T x_j\right) = \frac{2}{n}\sum_{i=1}^{n} x_i^T x_i.$$

- The sum of squared distances between points in the full space is

$$\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}^2 = n\sum_{i=1}^{n} x_i^T x_i = n\text{trace}(B) = n\sum_{i=1}^{n}\lambda_i.$$

- A measure of how to choose m is $P^{(m)} = \frac{\sum_{i=1}^{m}\lambda_i}{\sum_{i=1}^{n}\lambda_i}$.

2.5 A practical algorithm for classical MDS

- Obtain dissimilarities $\{\delta_{ij}\}$.
- Find matrix $A = [-\frac{1}{2}\delta_{ij}^2]$.
- Find matrix $B = [a_{ij} - a_{i.} - a_{.j} + a_{..}]$.
- Find the eigenvalues $\lambda_1, \ldots, \lambda_n$ and associated eigenvectors $v_1, \ldots, v_n$, where the eigenvectors are normalized so that $v_i^T v_i = \lambda_i$. If B is not PSD,let $\delta'_{ij} = \delta_{ij} + c(1 - \delta^{ij})$ and return to step 2.
- Choose an appropriate number of dimensions m (possibly use $P^{(m)}$ for a measure to choose m).
- The coordinates of the n points in the m dimensional Euclidean spaces are given by $x_{ij} = v_{ji}$ for $i = 1, \ldots, n, j = 1, \ldots, m$.