• classical multidimensional scaling

- assume that the observed \( n \times n \) proximity matrix \( D \) is a matrix of Euclidean distances derived from a raw \( n \times q \) data matrix, \( X \), which is not observed.

- define an \( n \times n \) matrix \( B \)

\[
PAP' = B = XX' = (X'M)(X'M)^T \quad M: \text{an orthogonal matrix}
\]

- the elements of \( B \) are given by

\[
b_{ij} = \sum_{k=1}^{q} x_{ik}x_{jk}
\]

- the squared Euclidean distances between the rows of \( X \) can be written in terms of the elements of \( B \) as

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

- idea: If the \( b_{ij} \)'s could be found in terms of the \( d_{ij} \)'s in the equation above, then we can derive \( X \) from \( B \) by factoring \( B \).

- to obtain \( B \) from \( D \), no unique solution exists unless a location constraint is introduced. Usually, the center of the columns of \( X \) are set at origin, i.e.,

\[
\sum_{i=1}^{n} x_{ik} = 0, \quad \text{for all } k
\]

- these constraints imply that sum of the terms in any row of \( B \) must be 0, i.e.,

\[
\sum_{j=1}^{n} b_{ij} = \sum_{j=1}^{n} \sum_{k=1}^{q} x_{ik}x_{jk} = \sum_{k=1}^{q} x_{ik} \left( \sum_{j=1}^{n} x_{jk} \right) = 0
\]

- Let \( T \) be the trace of \( B \). To obtain \( B \) from \( D \), notice that

\[
\sum_{i=1}^{n} d^2_{ij} = T + nb_{jj} \quad \sum_{j=1}^{n} d^2_{ij} = nb_{ii} + T
\]

\[
\sum_{i=1}^{n} \sum_{j=1}^{n} d^2_{ij} = 2nT
\]

- the elements of \( B \) can be found from \( D \) as

\[
b_{ij} = -\frac{1}{2} \left[ (d^2_{ij} - d^2_{i} - d^2_{j} + d^2_{..}) \right]
\]

where \( d^2_{i} = (\sum_{j=1}^{n} d^2_{ij})/n \), \( d^2_{j} = (\sum_{i=1}^{n} d^2_{ij})/n \), \( d^2_{..} = (\sum_{i=1}^{n} \sum_{j=1}^{n} d^2_{ij})/n^2 \)

- \( B \) can be written as

\[
B = V\Lambda V', \quad \text{spectral decomposition}
\]

where \( \Lambda = \text{diag}[\lambda_1, \ldots, \lambda_n] \) (\( \lambda_1 \geq \cdots \geq \lambda_n \)) is the diagonal matrix of eigenvalues of \( B \) and \( V = [V_1, \ldots, V_n] \) is the corresponding matrix of normalized eigenvectors (i.e., \( V_i'V_i = 1 \))

- Note: when \( D \) arises from an \( n \times q \) data matrix, the rank of \( B \) is \( q \) (i.e, the last \( n-q \) eigenvalues should be zero)

- So, \( B \) can be chosen as

\[
B = V^*\Lambda^* V^{+}, \quad V^{+} = V^* \Lambda^{1/2}
\]

where \( V^* \) contains the first \( q \) eigenvectors and \( \Lambda^* \) the first \( q \) eigenvalues

- Thus, a solution of \( X \) is \( X = V^*\Lambda^{1/2} \)
The adequacy of the \( q \)-dimensional representation can be judged by the size of the criterion

\[
\frac{(\sum_{i=1}^{q} \lambda_i)}{(\sum_{i=1}^{n} \lambda_i)}
\]

When the observed proximity matrix is not Euclidean, the matrix \( B \) is not positive-definite. In such case, some of the eigenvalues of \( B \) will be negative; correspondingly, some coordinate values will be complex numbers.

- If \( B \) has only a small number of small negative eigenvalues, it’s still possible to use the eigenvectors associated with the \( q \) largest positive eigenvalues.

- The adequacy of the resulting solution might be assessed using

  \[
  (\sum_{i=1}^{q} |\lambda_i|)/(\sum_{i=1}^{n} |\lambda_i|) \equiv p^{(1)}
  \]

  \[
  (\sum_{i=1}^{q} \lambda_i^2)/(\sum_{i=1}^{n} \lambda_i^2) \equiv p^{(2)}
  \]

- Some other issues

  - Metric scaling (define loss function for \( D \) and the distance matrix based on \( X \), called \textit{stress}, and find \( X \) to minimize stress)
  - Non-metric scaling (applied when the actual values of \( D \) is not reliable, but their orders can be trusted)
  - 3-way multidimensional scaling (proximity matrix result from individual assessments of dissimilarity and more than one individual is sampled)
  - Asymmetric proximity matrix (\( d_{ij} \neq d_{ji} \))

\[\textbf{Reading:}\] Reference, 5.1, 5.2