HUDM5124: Introduction to Multidimensional Scaling, Clustering, and Related Methods

Session 3: Introduction to metric MDS

Metric MDS

PROBLEM OF METRIC MDS:

GIVEN: a matrix Δ of "distance-like" proximities among *n* objects (the "dissimilarities"), possibly transformed so as to satisfy the metric axioms.

CONSTRUCT: an *n* x *r* matrix X (= the configuration of *n* points in a geometric space of r dimensions), such that the distances in the geometric space, D, (= the "model distances") are linearly related to the dissimilarities

(Note – the problem of determining the dimensionality *r* of the solution space must be faced. It may be determined by the data (= # of nonzero roots – see below), but often it is specified by the user based on interpretability.)

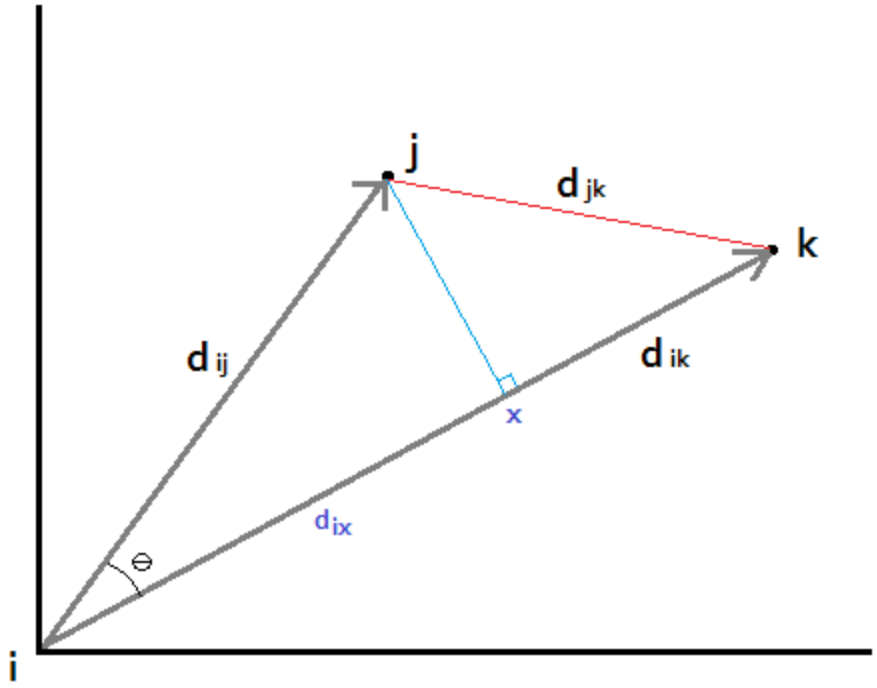
Background: Points as vectors, interpoint distances, and the Cosine Law

Represent two points by vectors (from the origin i):

$$\underline{k} = \vec{ik} = (x_{k1}, x_{k2}) \qquad \underline{j} = \vec{ij} = (x_{j1}, x_{j2})$$

Then the *scalar product* (dot product, inner product) of the two vectors is defined as:

$$j \cdot k = (x_{j1})(x_{k1}) + (x_{j2})(x_{k2})$$



Geometric interpretation:

$$j \cdot k = \left\|\vec{ij}\right\| \left\|\vec{ik}\right\| \cos(\theta)$$

i.e., *j·k* can be interpreted as the length of vector *ij* times the length of vector *ik* times the cosine of angle ɵ. Since cos(ɵ) = ||ix|| / ||ij||,

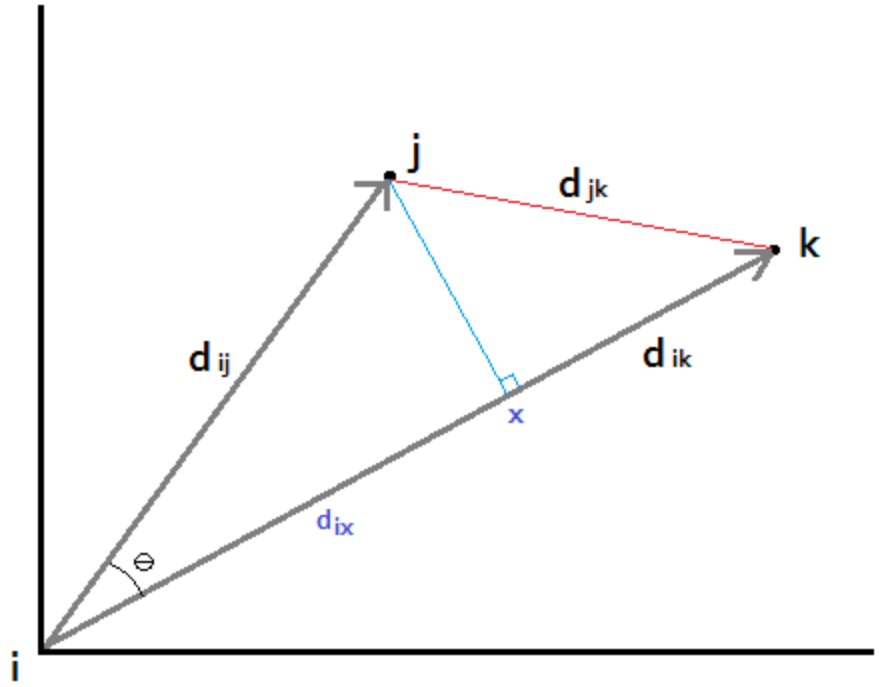
$$k \cdot j = \left\|\vec{ik}\right\| \left\|\vec{ix}\right\|$$

the Cosine Law

Switching notation from ||ik|| → d_{ik} we have:

$$k \cdot j = \|\vec{ik}\| \, \|\vec{ij}\| \cos(\theta)$$

$$= (d_{ik})(d_{ij}) \cos(\theta)$$



If k is assumed to lie along the X1 axis, it is easy to derive the Cosine Law:

Cosine Law: $$d_{jk}^2 = d_{ij}^2 + d_{ik}^2 - 2 d_{ij} d_{ik} \cos(\Theta)$$

→ $$j \cdot k = d_{ij} d_{ik} \cos(\Theta) = -\frac{1}{2}[d_{jk}^2 - d_{ij}^2 - d_{ik}^2]$$

This equation relates scalar products (which can be factored when represented in a matrix) to (squared) interpoint distances (d_{jk}) in a model space. [The (transformed) data may be thought of as estimates of these model space distances.]

Young-Householder (1938) Theorems

- If a square symmetric *n x n* matrix B_i of scalar products among stimulus vectors (defined w.r.t origin *i*) is *positive semidefinite*, then the "data distances" among the stimuli may be considered as the distances among points in a real, Euclidean space.
- The *rank* of the positive semidefinite matrix B_i is equal to the dimension of the set of points.
- Any positive semidefinite matrix B_i may be factored to obtain a matrix X, where $B_i = XX'$
- If the rank of B_i is r, where $r \leq n-1$, the matrix X is an $(n-1)$ x *r* rectangular matrix whose elements are the projections of the points on *r* orthogonal axes with origin at the point *i* (i.e. matrix X gives the coordinates of the points)

Torgerson's (1957) metric MDS algorithm

1) Using reasonable (and minimal) transformations, convert the proximities into distance-like numbers satisfying the metric axioms → Δ (n x n)
2) Square these "pseudo-distances" and double-center:
   → B* (= *n* x *n* matrix of "quasi-scalar products")

$$b^{*}_{ij} = -\frac{1}{2}\left[\delta^{2}_{ij} - \frac{1}{n}\sum_{i=1}^{n} \delta^{2}_{ij} - \frac{1}{n}\sum_{j=1}^{n} \delta^{2}_{ij} + \frac{1}{n^{2}}\sum_{i,\,j=1,1}^{n,n} \delta^{2}_{ij} \right]$$

3) Factor the B* matrix, extract *k* "important" (positive) roots, then the principal component loadings may be considered to be the coordinates (X) of the object points in *k*-space
4) Plot the *n* points in *k* dimensions (& interpret)

Assessing Fit of the Solution

- In metric MDS, we assume a linear relationship between the proximities and the distances in the "true" space.
- Therefore, the Pearson correlation between the dissimilarities and the derived model distances can be used to assess the goodness of fit of the derived solution.
- Model (Euclidean) distances are computed from the derived stimulus configuration (the final X matrix).

Class example:

model distances: dist = 3.000 5.003 6.001 4.004 5.001 1.041

original dissimilarities: prox = 2 4 3 5 4 0

> cor(prox,dist)

r = 0.749

this degree of linear fit is only moderate. The RSQ = 0.56, so about 56% of the variance in the proximities can be accounted for by the 2-dim solution

The additive constant problem

- When we transform the original proximities into "pseudo-distances", we may add a constant *c* to all entries of the dissimilarity matrix, so that they satisfy the metric axioms. This is "permissible" as a data analysis technique, since in metric MDS we usually assume that the dissimilarities are interval-level data (not ratio).
- We generally add the *minimum* constant necessary to satisfy the triangle inequality. This is based on the "linear subsets" rule: if three points lie in a straight line in the psychological space, then the distances among them should exactly satisfy the triangle inequality.
- But this may not be the optimal constant to add (esp. when *m* is small) because there may not be any linear subset, so one simple approach is to add a slightly larger constant.

The additive constant problem (cont.)

Why does this matter?

- If we add *too small* an additive constant (or none), then the data dissimilarities may not satisfy the metric axioms, thus we are trying to model a matrix that cannot be understood as distances
- As a result, the matrix B* derived from these distances may not be positive semidefinite, i.e. it may have negative eigenvalues.
- Using a larger additive constant may fix this problem, because adding a positive constant to all entries of a square symmetric matrix will increase the *trace* of the matrix, which will increase the size of the roots (eigenvalues).
- But this means that adding *too large* an additive constant might spuriously increase the apparent dimensionality.

The additive constant problem (cont.)

As can be seen in the demonstration below (using the class example data), adding *too small* an additive constant ($0<c<1.0$) to the dissimilarities can result in negative eigenvalues. But adding *too large* a constant ($c \geq 1.5$) can spuriously increase the apparent dimensionality (in this case, from 2 to 3).

Note that the value of c that exactly satisfies the TI for this example is 1. This may be an underestimate, if no three points in the "true" configuration fall on a line). According to Cailliez (1983), we should add the minimum constant that makes all eigenvalues non-negative, which is approximately 1.5 here. (Cailliez gives an analytic solution to compute this constant).

Table: Derived eigenvalues for Torgerson MDS solution, by magnitude of additive constant.

	Additive constant, c						
comp.	0	0.5	1	1.5	2	2.5	3
[1,]	16.566	20.250	24.190	28.385	32.833	37.533	42.484
[2,]	1.434	2.519	3.853	5.435	7.267	9.348	11.679
[3,]	0.000	0.000	0.000	0.555	1.400	2.495	3.838
[4,]	-0.500	-0.394	-0.043	0.000	0.000	0.000	0.000

An analytic solution to the additive constant problem

- Cailliez, F. (1983) The analytical solution of the additive constant problem. *Psychometrika,* **48**, 343–349.

- (See also Mardia, K. V. 1978. Some properties of classical multi-dimensional scaling. *Communications in Statistics – Theory and Methods, 7(13),* 1233–1241.)

SUMMARY

- Torgerson's "classical" MDS is based on geometric theorems relating scalar products between vectors to distances among pairs of points
- Thus, in a metric MDS solution (unlike in a plot of factor loadings from PCA or FA) we can directly interpret the distances among points in the derived solution (model) space
- Note that newer algorithms have been proposed to fit the metric MDS model (where distances in the solution space are assumed to be related to the proximity data by a linear transform)
- Currently, the state-of-the-algorithm for metric MDS is based on the numerical method of majorization, as implemented in programs such as SMACOF (R) or PROXSCAL (SPSS) (de Leeuw 1977, 1988; Groenen, Mathar & Heiser, 1995)