Distance Methodes & Multidimensional Scaling

Roderick Quiel

What is Multidimensional Scaling (MDS)

- A visual representation of distances or (dis)similarities between sets of objects.
- The primary objective is: to “fit” the original data into a low-dimensional coordinate system such that any “distortion” caused by the reduction in dimensionality is minimized.
- “Distortion”: generally refers to similarities or dissimilarities (distances) among the original data points.

Main approaches to MDS

- **Classical/Metric MDS** (quantitative data): Uses the actual magnitude of the original similarities.
- **Non-Metric MDS** (non-quantitative data): Given *N* items, uses only rank orders of the *N(N-1)/2* original similarities and not their magnitudes.

Distance, Similarities, and Disimilarities

- ***Dissimilarities*** represent the distance between two objects. This can be measured directly, as in the distance between two states or countries, or approximated. MDS algorithms use dissimilarities directly.

- ***Similarities*** represent how close (in some sense) two objects are. Assuming no ties, similarities are arranged in a strictly ascending order.

$$s_{i_1k_1} < s_{i_2k_2} < \cdots < s_{i_M,k_M}$$

Here $M = N(N-1)/2$ and $s_{i_1k_1}$ is the smallest of the M similarities.

Distance, Similarities, and Disimilarities

- Similarities can be converted to dissimilarities using the formula:

$$d_{ik} = \sqrt{s_{ii} + s_{kk} - 2s_{ik}}$$

where d_{ik} *represents a dissimilarity and* s_{ik} *represents a similarity.*

Classical/Metric MDS

- MDS takes as input an *(n x n) dissimilarity (distance) matrix D* containing the pairwise dissimilarities between all *n* data points.

- The matrix *D* must be symmetric and satisfy:

$$d_{ii} = 0, \;\; d_{ik} \geq 0, \;\;\; d_{ik} = d_{ki}$$

- The elements of *D* are denoted as:

$$d_{ik} = \sqrt{\sum_{j=1}^{p} \left(x_{ij} - x_{jk}\right)^2}$$

Classical/Metric MDS

- Given a dissimilarity (distance) matrix D, MDS seeks to find $\mathbf{X} = \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \in \mathbb{R}^p$ such that

$$d_{ik} \approx \ \| \mathbf{x}_i - \mathbf{x}_k \| \text{ as close as possible.}$$

- In classical MDS, we can get $d_{ik} = \ \| \mathbf{x}_i - \mathbf{x}_k \|$ if we multiply D^2 by $H = \mathbb{I}_n - \frac{1}{n} 1_n 1_n^{'}$ on both sides and $-\frac{1}{2}$. Let B be this new matrix.
- A solution $\mathbf{X}$ is then given by the eigendecomposition of $B = Q \Lambda Q^{'}$, that is $\mathbf{X} = \Lambda^{\frac{1}{2}} Q^{'}$.

Classical/Metric MDS

MDS Algorithm

1. Find matrix D and calculate the matrix $A = -\frac{1}{2}D^2$

2. From A calculate matrix B.

3. Then we have to find the largest p eigenvalues $\lambda_1 \geq \ldots \geq \lambda_p$ of B and their corresponding eigenvectors $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_p)$ s.t. $\mathbf{x}_i^{'}\mathbf{x}_i = \lambda_i$

4. A p-dimensional spatial configuration of the N items/objects is derived from the rows of $\mathbf{X}$.

Classical/Metric MDS

How to choose dimensions in classical MDS?

- We definitely want a small number of dimensions in order to have practical interpretations.
- Can be given by the rank of B or the number of nonzero eigenvalues.
- If B is positive semidefinite, then the number of nonzero eigenvalues gives the number of eigenvalues required for representing the distances d_{ik}.
- Calculation of the proportion of variation by p dimensions is given by:

$$P = \frac{\sum_{i=1}^{p} \lambda_i}{\sum_{i=1}^{n-1} \left| \lambda_i \right|},$$

P < 0.8 is a reasonable fit.

Classical/Metric MDS

Distance Scaling

- Unlike classical MDS, Metric MDS is an optimization process minimizing the stress function and is solved by iterative algorithms.

- Given a (small) dimension p and a monotone function f, Metric MDS seeks to find an optimal data matrix $\mathbf{X} \subset \mathbb{R}^p$ s.t.

$$f(d_{ik}) \approx \hat{d}_{ik} = \| \mathbf{x}_i - \mathbf{x}_k \|$$, as close as possible.

- This is now explicitly stated by

$$Stress = \left(\frac{\sum (d_{ik} - \hat{d}_{ik})^2}{\sum d_{ik}^2} \right)^{\frac{1}{2}}$$

Non-Metric MDS