Multi-Dimensional Scaling

Stephen Vardeman

Analytics Iowa LLC

ISU Statistics and IMSE

Objective

This material begins (as in clustering) with dissimilarities among N items, d_{ij}, that might be collected in an $N \times N$ proximity matrix $\mathbf{D} = (d_{ij})$. (These might, but do not necessarily, come from Euclidean distances among N data vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ in $\Re^p$.) The object of multi-dimensional scaling is to (to the extent possible) represent the N items as points $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N$ in $\Re^k$ with

$$\|\mathbf{z}_i - \mathbf{z}_j\| \approx d_{ij}$$

This is phrased precisely in terms of minimization of one of several possible "stress functions" $S(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N)$.

Kruskal-Shepard and Sammon criteria

The least squares (or Kruskal-Shepard) stress function (optimization criterion) is

$$S_{\mathrm{LS}}\left(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N\right) = \sum_{i<j} \left(d_{ij} - \|\mathbf{z}_i - \mathbf{z}_j\|\right)^2$$

This criterion treats errors in reproducing big dissimilarities exactly like it treats errors in reproducing small ones. A different point of view would make faithfulness to small dissimilarities more important than the exact reproduction of big ones. The so-called Sammon mapping criterion

$$S_{\mathrm{SM}}\left(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N\right) = \sum_{i<j} \frac{\left(d_{ij} - \|\mathbf{z}_i - \mathbf{z}_j\|\right)^2}{d_{ij}}$$

reflects this point of view.

A "local" version of MDS

Another approach to MDS that emphasizes the importance of small dissimilarities might be termed "local multi-dimensional scaling." Here one begins for fixed K with the symmetric set of index pairs

$$\mathcal{N}_K = \left\{ (i,j) \,\middle|\, \begin{array}{c} \text{the number of } j' \text{ with } d_{ij'} < d_{ij} \text{ is less than } K \\ \text{or the number of } i' \text{ with } d_{i'j} < d_{ij} \text{ is less than } K \end{array} \right\}$$

(an index pair is in the set if one of the items is in the K-nearest neighbor neighborhood of the other). A stress function that emphasizes the matching of small dissimilarities and not large ones is (for some choice of $\tau > 0$)

$$S_{\mathrm{L}}(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N) = \sum_{i<j \text{ and } (i,j)\in\mathcal{N}_K} (d_{ij} - \|\mathbf{z}_i - \mathbf{z}_j\|)^2 - \tau \sum_{i<j \text{ and } (i,j)\notin\mathcal{N}_K} \|\mathbf{z}_i - \mathbf{z}_j\|$$

A non-metric stress function

The preceeding are all "metric" scaling criteria in that the distances $\|\mathbf{z}_i - \mathbf{z}_j\|$ are meant to approximate the d_{ij} directly. An alternative is to use a non-metric stress function like

$$S_{\mathrm{NM}}(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N) = \frac{\sum_{i<j} \left(\theta(d_{ij}) - \|\mathbf{z}_i - \mathbf{z}_j\|\right)^2}{\sum_{i<j} \|\mathbf{z}_i - \mathbf{z}_j\|^2}$$

for vectors $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N$ and increasing functions $\theta(\cdot)$. $\theta(\cdot)$ that preserve/enforce the natural ordering of dissimilarities without attaching importance to their precise values. Iterative algorithms for optimization of this stress function alternate between isotonic regression to choose $\theta(\cdot)$ and gradient descent to choose the $\mathbf{z}_i$.

In general, if one can produce a small value of stress in MDS, one has discovered a k-dimensional representation of N items, and for small k, this is a form of "simple structure."