Correlation and Factor Analytic Approach to Multidimensional Scaling

Roberto N. Padua[1]

Date Submitted: September 2011
Date Revised: October 2011
Word Count: 1,539

Abstract

Multidimensional scaling (MDS) attempts to represent higher dimensional p-variate vectors in lower dimensional spaces such that the interitem proximities are closely preserved. The paper suggests two (2) new procedures for performing MDS that minimizes a stress function using pairwise correlations and by utilizing aspects of factor scores in factor analysis. Results reveal that the two (2) procedures perform as well as, if not better than, the classical minimization search algorithm.

Keywords: *multidimensional scaling, correlations, factor scores*

Introduction

Multidimensional scaling (MDS) aims to represent p-dimensional random vectors as lower q-dimensional vectors (q < p) in such a way that the original interitem proximities in R^p are as closely preserved as possible in R^q. Kruskal (1964) and Shepard (1980) pioneered in MDS development and proposed an algorithm that minimizes:

$$(1)\ldots Stress = \left[\frac{\sum\sum_{i<j}(d_{i,j}^{(p)} - d_{i,j}^{(q)})^2}{\sum\sum_{i<j}(d_{i,j}^{(p)})^2}\right]^{1/2}$$

where $d_{i,j}^{(p)}$ is the distance of item i from item j in R^p and $d_{i,j}^{(q)}$ the corresponding distance in R^q. Takane (1987) introduced a more popular stress measure:

$$(2)\ldots Stress = \left[\frac{\sum\sum_{i<j}({d_{i,k}^{(p)}}^2 - {d_{i,k}^{(q)}}^2)^2}{\sum\sum_{i<j}({d_{i,k}^{(p)}})^4}\right]^{1/2}$$

which lies between 0 and 1. In both instances, however, a search algorithm (usually, a steepest descent method) is implemented to look for points in R^q that minimize (1) or (2).

Applications of MDS abound in practice. A poverty mapping application was tried out by Miller (2001), Chen (2007) and others using MDS in q = 2 dimensions. Young

[1] Research Consultant, Jose Rizal Memorial State University

and Hammer (1987) provides an excellent account of the history and various applications of MDS in the social sciences, economics, engineering and the applied sciences.

The main problem with current MDS methodology is the difficulty in implementing the search algorithm. The present paper aims to address the following issue: Find the lowest dimension q $(X \epsilon R^q)$ to represent a p-dimensional vector Y $(Y \epsilon R^p)$ such that the interitem Euclidean distances in R^p are preserved. Corollarily, the second approach suggested in this paper addresses the more general MDS problem.

Maximal Dimension Reduction: The Correlation Approach

Let $X^{(p)} \epsilon R^p$ be a p-dimensional random vector with distribution function F(.) assumed absolutely continuous with respect to a Lebesgue measure. Let $X_1^{(p)}, X_2^{(p)}, \dots, X_n^{(p)}$ be iid F(.) and suppose that $X^{(p)}$ has finite second moment. i.e. each component X_i of $X^{(p)}$ has $E(X_i^2) < \infty$.
Let:

$$(3) \quad X_j^{(p)} = \begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{pj} \end{pmatrix} \quad , \quad j = 1, 2, \dots, n.$$

With respect to a Euclidean distance, it is possible to order the distances

$$d_{i,j}^{(p)} = \left[\sum_{k=1}^{p} \left(X_{kj} - X_{ki}\right)^2\right]^{\frac{1}{2}} \quad i \neq j, \text{ as:}$$

$$(4) \quad d_{i_1,j_1}^{(p)} \geq d_{i_2,j_2}^{(p)} \geq \cdots \geq d_{i_n,j_n}^{(p)}$$

where items i_1 and j_1 are the "farthest" items from each other. Without loss of generality, we assume that the component X_i of $X^{(p)}$ are standardized so that $E(X_i) = 0, var(X_i) = 1$, and the relation in (4) applies to the random variables.

To motivate the proposed approach for maximal dimension reduction, consider the bivariate observations:

$$X_1 = \begin{pmatrix} x_{11} \\ x_{21} \end{pmatrix}, X_2 = \begin{pmatrix} x_{12} \\ x_{22} \end{pmatrix}, \dots, X_n = \begin{pmatrix} x_{1n} \\ x_{2n} \end{pmatrix}.$$

If $d(X_1, X_2) < d(X_3, X_4)$ then:

$$(5) \quad (X_{11} - X_{12})^2 + (X_{21} - X_{22})^2 < (X_{13} - X_{14})^2 + (X_{23} - X_{24})^2.$$

We wish to reduce the dimension from p = 2 to q = 1 while preserving $d\left(X_1^{(q)}, X_2^{(q)}\right) < d\left(X_3^{(q)}, X_4^{(q)}\right)$. Suppose that $corr\left(X_{1j}, X_{2j}\right) = 1$, then this means that we can express the second coordinate as a linear multiple of the first coordinate, viz. $\left(X_{2j} = KX_{1j}\right)$. Using this relation and substituting into (5), we obtain:

$$(6)\quad (K^2+1)(X_{11}-X_{12})^2 < (K^2+1)(X_{13}-X_{14})^2$$

Or: $$(X_{11}-X_{12})^2 < (X_{13}-X_{14})^2$$

$$d\left(X_1^{(q)}, X_2^{(q)}\right) < d\left(X_3^{(q)}, X_4^{(q)}\right).$$

Inequality (6) says that we can drop the second coordinate (hence, reduce the dimension to q = 1) and still preserve the inequality.

In general, the same reasoning follows if $corr\left(X_{1j}, X_{2j}\right) = \phi \neq 0$. Express $X_{2j} = \phi X_{1j}$ and note that $\left(X_{1j}, X_{2j}\right) = E\left(\phi {X_{1j}}^2\right) = \phi\, E\left({X_{1j}}^2\right) = \phi$. What remains to be investigated is the case when the component are orthogonal, that is, if $corr\left(X_{1j}, X_{2j}\right) = 0$. In this case, no linear relationship exists between $\left(X_{1j} and\ X_{2j}\right)$ and we conclude that no reduction is possible that preserves (5). We state these observations as a Theorem:

Theorem: *Let $X^{(p)}$ be a p-dimensional random vector such that $\rho_{ij} \neq 0,\ i \neq j$. Let γ be the number of distinct pairs x_i and x_j for which $\rho_{ij} \neq 0$. Let γ be the number of distinct pairs x_i and x_j for which $\rho_{ij} \neq 0,\ i \neq j$. Then, the p-dimensional random vector $X^{(p)}$ can be reduced to a q-dimensional random vector $X^{(q)}$, $q < p, q = p - \gamma$ such that the original inter-item distance orderings are preserved* . ■

The proof is an exercise in algebra. In order to implement Theorem 1 in practice, we construct the $\frac{p(p-1)}{2}$ correlation matrix and test:

Ho:$\rho_{ij} = 0$ against Ha: $\rho_{ij} \neq 0$.

We are interested only in those ρ_{ij} for which the hypothesis in rejected.

Algorithm:

1. Standardized the components of $X_1^{(p)}, X_2^{(p)}, \dots, X_n^{(p)}$ by subtracting the means and dividing by their standard deviations.
2. Obtain the $M = \frac{n(n-1)}{2}$ inter-item distances d_{ij} and arrange:

$$d_{i_1,j_1}^{(p)} > d_{i_2,j_2}^{(p)} > \cdots > d_{i_n,j_n}^{(p)}$$

3. Compute the $Q = \frac{p(p-1)}{2}$ correlation coefficient $\hat{\rho}_{ij}$.
4. Test Ho: $\rho_{ij} = 0$ against Ha: $\rho_{ij} \neq 0$.
5. Collect all $\rho_{ij} \neq 0$ and use either X_i or X_j. Put the retained variables in $X_i^{(q)}$.
6. Obtain the inter-item distances in R^q and arrange:

$$d^{(q)}_{i_1,j_1} > d^{(q)}_{i_2,j_2} > \cdots > d^{(q)}_{i_n,j_n}$$

7. Compute the stress function.

The component of $X^{(q)}$ obtained in this manner are now mutually orthogonal and cannot be reduced further by the correlation approach. However, if it is desired to reduce the dimension further, then we suggest using the classical MDS search algorithm.

In the correlation approach, we drop one of the variables X_i or X_j when $corr(X_i, X_j) = \rho_{ij} > 0$. The vector $X^{(q)}$ then amounts to a projection of the vector $X^{(p)}$ in R^q. The correlation approach ensures preservation of the rank ordering but does not guarantee that the stress function will be minimum there.

Let $V_1, V_2, V_3, V_4 \epsilon R^p$ such that $d(V_1, V_2) < d(V_3, V_4)$ and let $V_1^*, V_2^*, V_3^*, V_4^* \epsilon R^q$ obtained by the correlation approach. Suppose that $V_1 = (X_1, X_2, X_3, \ldots, V_p)'$ and that $corr(X_i, X_j) = \rho_{ij} > 0$. Instead of dropping either X_i or X_j, we seek a linear combination $X_j + \varphi X_i$ such that $d(V_1^*, V_2^*) < d(V_3^*, V_4^*)$ and:

$$(7)\ \gamma = d^2(V_3^*, V_4^*) - d^2(V_1^*, V_2^*) \quad \text{is minimum, } \gamma \geq 0.$$

The problem reduces to a one-dimensional minimization problem. The solution to (7) minimizes the Takane (1987) stress function.

Factor Score Representation

A generalization of the correlation approach in Section 2 consists of combining in some meaningful way, all those variables that highly correlate with one another. The linear combination of these highly correlated variables will then represent one (1) dimension. A natural way to group variables according to their correlations with each other is through factor analysis.

The usual orthogonal factor model is:

$$(8)\ X = LF + \varepsilon$$

Where X is a p x 1 random vector (observed), L is a p x q matrix of factor loadings, F is an unknown q x 1 vector of factors, ε is a random error with $E(\varepsilon) = 0$ and $Var(\varepsilon) = \psi$ a diagonal matrix. The model assumes $E(F) = 0, Var(F) = I \; and \; cov(F, \varepsilon) = 0$.

In the extreme case when $cov(X) = \Sigma$ is diagonal, i.e. the components of X are uncorrelated, then the individual x_i's themselves are the factors and no dimension reduction is possible.

The estimated factor scores are:

$$(9)\ \hat{f}_j = l_{i1}X_1 + l_{i2}X_2 + \cdots + l_{ip}X_p, \quad i = 1, 2, \ldots, q$$

and we propose using only those $l_{ij} > 0$. This latter restriction stems from our desire to combine highly correlated random variables only.

Numerical Illustration and Simulation

We provide one (1) numerical illustration of the procedures and carried out extensive Monte Carlo simulation to assess the performance of the proposed procedures.

Numerical Illustration

The original random vector consists of 4 dimensions (height, weight, IQ, Math). Five (5) individuals of various heights, weights, IQ's and Mathematical abilities are illustrated. Table 1 shows the original data and their standardized values.

Table 1 Original Data and Standardized Values

Original Data					Standardized Data			
Individual	Height	Weight	IQ	Math	Height	Weight	IQ	Math
A	5.5	72	110	80	-0.447	-0.525	-1.412	-1.372
B	5.3	70	120	85	-1.341	-1.180	-0.581	-.610
C	5.7	75	130	90	0.447	0.459	0.249	0.153
D	5.9	78	135	94	1.341	1.443	0.664	0.762
E	5.6	73	140	96	0.000	-0.197	1.079	1.067

The computed correlations are:

$corr(H, W) = 0.990\ (p < .001)$
$corr(IQ, Math) = 0.997\ (p < .001)$
$corr(IQ, H) = 0.604\ (p = 0.281)$
$corr(IQ, W) = 0.574\ (p = 0.314)$
$corr(MAth, H) = 0.631\ (p = 0.254)$
$corr(MAth, W) = 0.600\ (p = 0.285)$

so that only two (2) correlations are found significant beyond the .001 level. Thus, it appears that the dimension could be effectively reduced to q = 2. Table 2 shows d(A, B), d(B, C), d(C, D) and d(D, E) from among the ten (10) possible distances.

Table 2 Sample Distances

Q=2 Reduced Space						
Distances	Standardized Data	Rank	(H, Math)	Rank	(H, IQ)	Rank
d(A, B)	1.581	2	1.175	2	1.220	2
d(B, C)	2.676	4	1.945	4	1.972	4
d(C, D)	1.521	1	1.083	1	0.986	1
d(D, E)	2.180	3	1.376	3	1.404	3

Tabular values shows that the rank orders of the original distances are perfectly preserved in the reduced q = 2 dimensional space. For these four (4) distances the Takane (1980) stress is:

Reduced to (H, Math): Stress = 0.84
Reduced to (H, IQ) : Stress = 0.82

Both of which are less than 1.

Table 3 shows two (2) factor score representation of the data viz - a - viz the distances:

Table 3 Two-Factor Representation

Two-Factor Representation						
Distances	Original	Rank	F1	F2	New Vector Distance	Rank
d(A, B)	1.581	2	-1.656	-0.580	1.321	2
d(B, C)	2.676	4	-0.708	-1.500	2.249	4
d(C, D)	1.521	1	0.239	0.540	1.274	1
d(D, E)	2.180	3	0.847	1.660	1.830	3
			1.277	-0.120		

Again the rank orders of the distances are exactly the same for p = 4 and q = 2 representations. The Takane stress is:

Stress = 0.29

which shows an excellent fit.

References

Kruskal, J.B. (1964). "Non-Metric Multidimensional Scaling: A Numerical Method" Psychometrika, 29, no. 1, pp. 115-129.

Shepard, R.N. (1980). "Multidimensional Scaling, Tree-Fitting and Clustering" Science, 210, no.4468, pp. 390-398.

Takane, Y., F.W. Young, and J.D. Leeuw (1977). "Non-Metric Individual Differences Multidimensional Scaling" Psychometrika, 42, pp. 7-67.

Young,F.W and R.M. Hamer (1987). Multidimensional Scaling: History, Theory and Applications. Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.