Multidimensional Scaling

Kaustav Paul Ananyo Dey Shreya Chatterjee

Instructor: **Dr. Swagata Nandi**

Date Last Edited: April 3, 2024

Contents

Introduction

Multidimensional scaling (MDS) is a technique that translates similarities (or differences) between pairs of items into distances within a lower-dimensional space. The data, for example, may be correlations among intelligence tests, and the MDS representation is a plane that shows the tests as points that are closer together the more positively the tests are correlated. By visually depicting these correlations, MDS allows analysts to visually inspect the data, often revealing patterns not apparent in raw numerical arrays. Additionally, MDS can be utilized to construct models for assessing dissimilarities. For example, given two objects of interest, one may explain their perceived dissimilarity as the result of a mental arithmetic that mimics the distance formula.According to this model, the mind generates an impression of dissimilarity by adding up the perceived differences of the two objects over their properties.

The basic purpose of our study is to demonstrate how multidimensional scaling can help systematize data in areas where organizing concepts and underlying distributions are not well developed. Multidimensional scaling is simply a useful mathematical tool that enables us to represent the similarities (or dissimilarities) of objects spatially as in a map.

Why MDS?

One of the principal advantages of MDS is its **ease of interpretation**. MDS translates the given data as points in low dimensional spaces keeping similarities and dissimilarities among the points in means of distances of the corresponding plotted points. Graphical representations always create clearer impression about the data structure than mere numerical summaries. Let us consider an example. In Table 1, we are provided with the mutual distances of ten Spanish cities. Suppose a random person has no access to the geographical map of the concerned region but he is provided with distance matrix, it will be easier for him to get an idea of the distances between the cities have the cities been represented as points with distances between the plotted points being similar to the original geographical distances as in Figure 2 . Hence MDS comes into great help in such a situation.

Let us consider another scenario. It may happen that instead of values of the concerned variables we are provided with some measure of similarity and dissimilarity between the objects. For instance, a tea taster upon tasting a number of tea flavors rates how much the taste of one flavor differs from another. In this case also we can use MDS proficiently.

Moreover MDS represents the objects as points in low dimensional spaces (mostly two or three dimensional spaces are taken in to consideration). Hence it can be used as a tool of dimension reduction.

MDS can also be used to find clusters among points. Suppose the data points can be classified into a number of groups, in that case the points coming from the same group can be expected to be similar to each other. So MDS will reflect clusters among the plotted points. Such examples will be considered later in the study.

Table 1: Matrix representation of the distance between eight Australian cities

	Adelaide	Alice Springs	Brisbane	Darwin	Hobart	Melbourne	Perth	Sydney
Adelaide	0	1328	1600	2616	1161	653	2130	1161
Alice Springs	1328	0	1962	1289	2463	1889	1991	2026
Brisbane	1600	1962	0	2846	1788	1374	3604	732
Darwin	2616	1289	2846	0	3734	3146	2652	3146
Hobart	1161	2463	1788	3734	0	598	3008	1057
Melbourne	653	1889	1374	3146	598	0	2720	713
Perth	2130	1991	3604	2652	3008	2720	0	3288
Sydney	1161	2026	732	3146	1057	713	3288	0

The Basic Setup

Suppose a set of n objects is under consideration and between each pair objects (r, s) there is a measurement δ_{rs} of the "dissimilarity" between objects. For example the set of objects might be ten bottles of whisky, each from a different distillery. The dissimilarity δ_{rs} might be an integer score between zero and ten given to the comparison of the r^{th} and the s^{th} whiskys by an expert judge of malt whisky. The judge would be given a tot from the r^{th} bottle and one from the s^{th} and then score the comparison: 0-the whiskies are so alike she/he cannot tell the difference, to 10-the whiskies are totally different. The judge is presented with all forty-five possible pairs of whiskies, and after a pleasant day's work, provides the data analyst with a total set of dissimilarities $\{\delta_{rs}\}$.

We may be provided with a different type of situation where we have values corresponding to p variables corresponding to the n objects. We represent

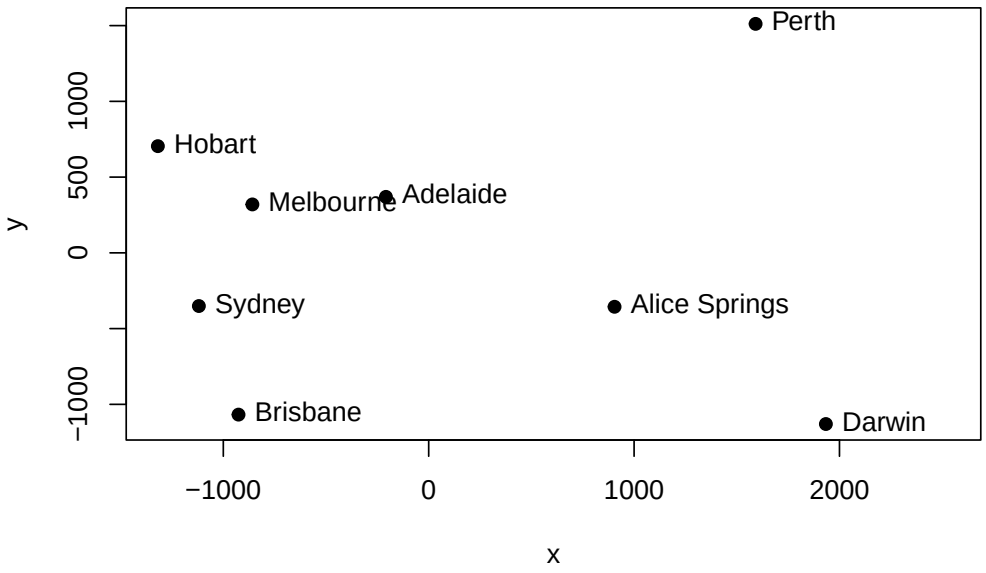

Figure 1: Representation of the cities as points on two dimensional space after MDS

the information in vector representation, as $\mathbf{x_1}$, $\mathbf{x_2}$, ..., $\mathbf{x_n}$ where each $\mathbf{x_1}$, $\mathbf{x_2}$, ..., $\mathbf{x_n}$ are p-vectors. In this case we have to obtain the mutual distance (dissimilarity) between the objects.

A narrow definition of multidimensional scaling (often abbreviated to MDS) is the search for a low dimensional space, usually Euclidean, in which points in the space represent the objects (whiskies), one point representing one object, and such that the distances between the points in the space, $\{\delta_{rs}\}$, match, as well as possible, the original dissimilarities $\{\delta_{rs}\}$. The techniques used for the search for the space and the associated configuration of points form metric and nonmetric multidimensional scaling.

Some Definitions

We describe preliminary material on proximity and metrics.

Proximity literally means nearness in space, time or in some other way. The "nearness" of objects, individuals, stimuli needs definition and measurement prior to statistical analysis. In some situations, this is straightforward, but in others, difficult and controversial. Measures of proximity are of two types: similarity and dissimilarity with the obvious interpretation of measuring how similar or dissimilar objects are to each other. In this section, we introduce some of the terminology and define some of the proximity measures discussed in the following chapters.

Definiton (Distance Matrix): An $(n \times n)$ matrix $\mathbf{D}$ is called a distance matrix if it is symmetric and

$$d_{rr} = 0,\ d_{rs} \geq 0, r \neq s.$$

The first property above is called reflectivity, and the second property is called non-negativity. Note that there is no need to satisfy the triangle inequality.

Definiton (Euclidean Distance): The Euclidean distance between two points $\mathbf{p} = (p_1, ..., p_n)$ and $\mathbf{q} = (q_1, ..., q_n)$ in $\mathbb{R}^n$ is given by the formula,

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2}$$

Definition (Euclidean Distance Matrix): A distance matrix $\mathbf{D}$ is called Euclidean if there exists a configuration of points in some Euclidean space

whose interpoint distances are given by $\mathbf{D}$; that is, if for some p, there exists points $\mathbf{x_1}, \ldots, \mathbf{x_n} \in \mathbb{R}^p$ such that}

$$d_{rs}^2 = (\mathbf{x_r} - \mathbf{x_s})^T(\mathbf{x_r} - \mathbf{x_s})$$

where ^{T} denotes the transpose of a matrix.

Definition (Metric): A function $d : X \times X \to \mathbb{R}$ is called a metric if the following conditions are fulfilled for all $x, y, z \in X$:

- (reflectivity) $d(x, x) = 0$
- (positivity) $d(x, y) > 0$ for $x \neq y$
- (symmetry) $d(x, y) = d(y, x)$
- (triangle inequality) $d(x, y) \leq d(x, z) + d(z, y)$

A metric space (X, d) is a set X equipped with a metric $d : X \times X \to \mathbb{R}$
In some situations we start not with distances between n objects, but with similarities.

Definition (Similarity Matrix): An $(n \times n)$ matrix $\mathbf{C}$ is called a similarity matrix if $c_{rs} = c_{sr}$ and if $c_{rs} \leq c_{rr}$ for all r, s
To use the techniques of the preceding method, it is necessary to transform the similarities to distances. A useful transformation is the following. The standard transformation from a similarity matrix $\mathbf{C}$ to a distance matrix $\mathbf{D}$ is defined by

$$d_{rs} = (c_{rr} - 2c_{rs} + c_{ss})^{\frac{1}{2}}$$

Hence, $\mathbf{D}$ is a distance matrix.
Similarities (dissimilarities) are constructed from a data matrix for the objects. These are then called similarity (dissimilarity) coefficients. Several authors for an example Cormack, Jardine and Sibson, Anderberg, Sneath and Sokal, Jackson *et al.*, Snijders *et al.* discuss various similarity and dissimilarity measures together with their associated problems. The following synthesis of the work of these authors attempts to outline the main ideas behind forming dissimilarities from a data matrix. Let, $\mathbf{X} = [x_{ri}]$ denoted the data matrix obtained for n objects on p variables. The vector of observations for the r^{th} object is denoted by $\mathbf{x_r} = [\mathbf{x_r^T}]$.

- *Quantative Data :* The following table gives some of possible dissimilarity measures for quantitative data that are in particular, continuous, possibly discrete, but not binary.

Euclidean Distance	$\delta_{rs} = \{\sum_i (x_{ri} - x_{si})^2\}^{\frac{1}{2}}$
Mahalanobis Distance	$\delta_{rs} = (x_r \text{ - } x_s)^T \Sigma^{-1} (x_r \text{ - } x_s)$
Minkowski metric	$\delta_{rs} = \{\sum_i w_i \lvert x_{ri} - x_{si}\rvert^{\lambda}\}^{\frac{1}{\lambda}}, \lambda \geq 0$
Bhattacharya distance	$\delta_{rs} = \{\sum_i (x_{ri}^{\frac{1}{2}} - x_{si}^{\frac{1}{2}})^2\}^{\frac{1}{2}}$
Bray-Curtis	$\delta_{rs} = \frac{1}{p} \frac{\sum_i \lvert x_{ri} - x_{si}\rvert}{\sum_i (x_{ri} + x_{si})}$

- *Binary Data :*

		Object s		Total
		1	0	
Object r	1	a	b	a+b
	0	c	d	c+d
Total		a+c	c+d	p = a+b+ c+d

The following table gives some of possible dissimilarity measures for binary data

Hamman	$s_{rs} = \frac{a-(b+c)+d}{a+b+c+d}$
Braun, Blanque	$s_{rs} = \frac{a}{max\{(a+b),(a+c)\}}$
Kulczynski	$s_{rs} = \frac{a}{(b+c)}$
Yule	$s_{rs} = \frac{ad-bc}{ad+bc}$
Russell, Rao	$s_{rs} = \frac{a}{(a+b+c+d)}$

Prime Goals

Suppose there are n objects with dissimilarities $\{\delta_{rs}\}$. Metric MDS attempts to find a set of points in a Euclidean space where each point represents one of the objects and the distance between points $\{\delta_{rs}\}$ are such that

$$d_{rs} \approx f(\delta_{rs}) \tag{1}$$

where f is a continuous parametric monotonic function. The function f can either be the identity function or a function that attempts to transform the dissimilarities to a distance-like form. Mathematically, let the objects comprise a set O. Let the dissimilarity, defined on $O \times O$, between objects r and s be $\delta_{rs}(r, s \in O)$. Let ϕ be an arbitrary mapping from O to E, where E is usually a Euclidean space, but not necessarily so, in which a set of points are to represent the objects. Thus let $\phi(r) = x_r (r \in O, x_r \in E)$, and let $X = \{x_r : r \in O\}$, the image set. Let the distance between the points $\mathbf{x_r}$, $\mathbf{x_s}$ in $\mathbf{X}$ be given by d_{rs}.

The aim is to find a mapping ϕ, for which d_{rs}. is approximately equal to $f(\delta_{rs})$ for all $r, s \in O$. The two main metric MDS methods, classical scaling and least squares scaling, will be considered here.

Nature of the Solution:

The configuration produced by any MDS method is indeterminate with respect to translation, rotation, and reflection. 1n the two-dimensional case of road distances the whole configuration of points can be "shifted" from one place in the plane to another and the whole configuration can be "rotated" or "reflected".

In general, if $P_1, ..., P_n$ with coordinates $\mathbf{x_i'} = (x_{i1}, x_{i2}, ..., x_{ip})$, $i = 1, ..., n$, represents an MDS solution in p dimensions, then

$$\mathbf{y_i} = \mathbf{A}\mathbf{x_i} + \mathbf{b}$$

is also a solution, where $\mathbf{A}$ is an orthogonal matrix and $\mathbf{b}$ is any vector. This is because multiplying a matrix by an orthogonal matrix only leads to rotation of the axes. Hence by the above matrix operation the solutions are just rotated and shifted from one position to another on the two dimensional plane which does not affect the feasibility to be a solution in an MDS problem.

Types of Solution:

Methods of solution using only tbe. rank order of the distances

$$d_{r_1 s_1} < d_{r_2 s_2} < ... < d_{r_m s_m}, \qquad m = \frac{1}{2} n(n-1),$$

where $(r_1, s_1), ..., (r_m, s_m)$ denote all pairs of subscripts of r and s, $r < s$, are titled as *non-metric methods of multidimensional scaling*.

The rank orders are invariant under monotone increasing transformations f of the d_{rs}, i.e.

$$d_{r_1 s_1} < d_{r_2 s_2} < ... \Leftrightarrow f(d_{r_1 s_1}) < f(d_{r_2 s_2}) < ...$$

Therefore the configurations which arise from non-metric scaling are indeterminate not only with respect to translation, rotation, and reflection, but also with respect to uniform expansion or contraction.

Solutions which try to obtain P, directly from the given distances are called *metric methods*. These methods derive P, such that, in some sense, the new

distances $\hat{d}_{rs}$ between points P_r, and P_s, are as close to the original d_{rs} as possible.

One important use of MDS is seriation. The aim here is to order a set of objects chronologically on the basis of dissimilarities or similarities between them. Suppose the points in the MDS configuration in $k = 2$ dimensions lie nearly on a smooth curve. This property then suggests that the differences in the data are in fact one-dimensional and the ordering of the points along this curve can be used to seriate the data.

Classical Scaling

Classical scaling originated in the 1930s when Young and Householder (1938) showed how starting with a matrix of distances between points in a Euclidean space, coordinates for the points can be found such that distances preserved. Torgerson (1952) brought the subject to popularity busing the technique for scaling.

Starting with a distance matrix D, the object of MDS is to find points $P_1, P_2, ..., P_n$ in k dimensions such that if $\hat{d_{rs}}$ denotes the Euclidean distance between P_r and P_s, then $\hat{\mathbf{D}}$ is "similar" in some sense to $\mathbf{D}$. The points P_r are unknown and usually the dimension k is also unknown. In practice, on usually limits the dimension to $k = 1$, 2 or 3 in order to facilitate the interpretation of the solution. The following theorem enables us to tell whether $\mathbf{D}$ is Euclidean, and, if so, how to find a corresponding configuration of points. For any distance matrix $\mathbf{D}$, let

$$\mathbf{A} = ((a_{rs})), \quad a_{rs} = -\frac{1}{2}d_{rs}^2 \tag{2}$$

and set $\mathbf{B} = \mathbf{HAH}$ where $\mathbf{H} = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T$ is the $(n \times n)$ centering matrix.

Theorem: Let $\mathbf{D}$ be a distance matrix and defined by $\mathbf{B}$ by previous, then $\mathbf{D}$ is Euclidean *if and only if* $\mathbf{B}$ is p.s.d. In particular, the following results hold:

1. If $\mathbf{D}$ is the matrix of Euclidean interpoint distances fo a configuration $\mathbf{Z} = (z_1, ..., z_n)^T$, then

$$\mathbf{b_{rs}} = (\mathbf{z_r} - \bar{\mathbf{z}})^T(\mathbf{z_s} - \bar{\mathbf{z}}), \qquad r, s = 1, ..., n$$

   In matrix form it becomes $\mathbf{B} = \mathbf{(HZ)(HZ)}^T$ so $\mathbf{B} \geq 0$. Note that $\mathbf{B}$ can be interpreted as the "*centred inner product matrix* for configuration of $\mathbf{Z}$

2. Conversely, if $\mathbf{B}$ is p.s.d of rank p then a configuration corresponding to $\mathbf{B}$ can be constructed as follows, let $\lambda_1 > ... > \lambda_p$ denote the positive eigenvalues of $\mathbf{B}$ with corresponding eigenvectors $\mathbf{X} = (\mathbf{x_{(1)}}, ..., \mathbf{x_{(p)}})^T$ normalized by

$$\mathbf{x_{(i)}}^T \mathbf{x_{(i)}} = \lambda_i, \qquad i = 1, 2, ..., p$$

Then the points P, in $\mathbb{R}^p$ with coordinates $\mathbf{x_r} = (\mathbf{x_{(r1)}}, ..., \mathbf{x_{(rp)}})^T$ (so x_r is the r^{th} row of $\mathbf{X}$) have interpoint distances given by $\mathbf{D}$. Further, this configuration has centre of gravity $\bar{x} = 0$, and $\mathbf{B}$ represents the inner product matrix for this configuration

To be of practical use, a configuration of points needs to be found for a set of dissimilarities $\{\delta_{rs}\}$ rather than simply for true Euclidean distances between points $\{d_{rs}\}$.

Suppose dissimilarities $\{\delta_{rs}\}$ are used instead of distances d_{rs} to define matrix $\mathbf{A}$, which is then doubly centred to produce matrix $\mathbf{B}$ as just described. Then it is interesting to ask under what circumstances $\mathbf{B}$ can give rise to a configuration of points in Euclidean space, using the spectral decomposition, so that the associated distances $\{\delta_{rs}\}$ are such that $d_{rs} = \delta_{rs}$ for all r, s. The answer is that if $\mathbf{B}$ is positive semi-definite pf rank p, then configuration in p dimensional Euclidean space can be found.

The next question to be asked is how many dimensions are required in general for the configuration of points produced from a positive semi-definite matrix $\mathbf{B}$ of dissimilarities. It can be shown that $\mathbf{B}$ has at least one zero eigenvalue, since $\mathbf{B1 = HAH1 = 0}$. Thus a configuration of points in an $n-1$ dimensional Euclidean space can always be found whose associated distances are equal to the dissimilarities $\{\delta_{rs}\}$. Choose the configuration in R^k whose coordinates are determined by the first k eigenvectors of $\mathbf{B}$. If the first k eigenvalues of $\mathbf{B}$ are "large" and positive, and hopefully, the interpoint distances of this configuration will closely approximate $\mathbf{D}$. This configuration is called the *Classical solution to the MDS problem in k dimensions*. It is metric solution.

Another suitable question can be what will happen if $\mathbf{B}$ is not positive semi-definite. Then a constant can be added to all the dissimilarities (except the self-dissimilarities δ_{rr}) which will then make $\mathbf{B}$ positive semi-definite. Thus forming new dissimilarities, $\{\delta_{rs}\prime\}$ as $\delta_{rs}\prime = \delta_{rs} + c(1 - \delta_{rs})$, where c is an appropriate constant and δ_{rs} the Kronecker delta will make $\mathbf{B}$ positive semi-definite. Once $\mathbf{B}$ has been made positive semi-definite, a Euclidean space can be found as before where distances d_{rs} are exactly equal to dissimilarities $\delta_{rs}\prime$.

A Practical Algoritm to Classical MDS

Suppose we are given a distance matrix D which we hope can approximately represent the inter-point distances of a configuration in a Euclidean space of low dimension k (usually $k = l, 2$, or 3). The matrix D mayor may not be Euclidean; however, even if D is Euclidean, the dimension of the space in which it can be represented will usually be too large to be of practical interest. Classical MDS uses the following algorithm to determine the required solution:

1. From D construct the matrix $A = -\frac{1}{2}d_{rs}^2$
2. Obtain the matrix B with elements $b_{rs} = a_{rs} - \bar{a}_{i\cdot} - \bar{a}_{\cdot s} + \bar{a}_{\cdot\cdot}$
3. Find the k largest eigenvalues $\lambda_1 > ... > \lambda_k$ of $B(k$ chosen ahead of time), with corresponding eigenvectors $X = (\mathbf{x_{(i)}}, ..., \mathbf{x_{(k)}})$ which are normalized by $\mathbf{x}'_{\mathbf{(i)}}\mathbf{x_{(i)}} = \lambda_i, i = 1, ..., k$. (We are supposing here that the first k eigen values are all positive.)
4. The required coordinates of the points P_r are $\mathbf{x_r} = (x_{r1}, ..., x_{rp})'$, $r = 1, ..., k$, the rows of X. r= 1. . . . , k, the rows of X.

We will now apply this algorithm over a data set.

An Example

We will now apply Classical MDS over Fisher's `iris` data set.

The data set contains measurements in centimeters of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of 3 species of iris. The species are *Iris setosa*, *versicolor*, and *virginica*. A glimpse of the data set is given below:

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa

We have used the `cmdscale` function in the `stats` package in `R` to perform the Classical MDS algorithm over the data points. The function returns the

optimal points to be plotted on a k dimensional space. We have considered $k = 2$ in this case. We have plotted the points in Figure 2.

Figure 2: Classical MDS on iris data for $k = 2$

From the plot we can get a clear idea about the dissimilarities in the features corresponding to different flowers. The flowers belonging to the same category are put close to each other indicating that the extent of dissimilarity between two flowers of same category is lesser than those between two belonging to different categories. Thus in this case MDS helps us in finding clusters among the data points.

Similarities

We may also be provided with similarities among the data points instead of distances (or dissimilarities). We have already defined a similarity matrix in a previous section. To use the techniques of the preceding sections, it is necessary to transform the similarities to distances. We have previously discussed a transformation to convert a similarity matrix to a distance matrix as follows,

$$d_{rs} = (c_{rr} - 2c_{rs} + c_{ss})^{\frac{1}{2}}$$

In fact the transformed distance is an Euclidean distance matrix.

Theorem: If $C \geq O$, then the distance matrix D defined by the above standard transformation is Euclidean, with centered inner product matrix $B = HCH$.

Classical Scaling and Principal Components

Suppose X is a data matrix of dimension $n \times p$. The sample covariance matrix obtained from X is $S = (n-1)^{-1}X^TX$, where it is assumed that the data have been mean corrected. Principal components are obtained by finding eigen values $\{\lambda_i : i = 1, ..., p\}$ and right eigen vectors $\{\mathbf{e_i} : i = 1, ..., p\}$ of S, and then the ith principal component is given by $u_i = \mathbf{e_i}^T\mathbf{x}, (i = 1, ..., p)$

Suppose on the other hand Euclidean distance is used on the data matrix X to define dissimilarities among the n dissimilarities among the n individuals or objects. The dissimilarities will be given by

$$\delta_{rs}^2 = (\mathbf{x_r} - \mathbf{x_s})^T(\mathbf{x_r} - \mathbf{x_s})$$

and hence when these dissimilarities are subjected to classical scaling, $b_{rs} = \mathbf{x_r}^T\mathbf{x_s}$ and $B = XX^T$.

As before, let the eigenvalues of B be $l_i (i = 1, ..., p)$ with associated eigen vectors $\mathbf{v_i} (i = 1, ..., n)$.

It is a well known result that the eigen values of XX^T are the same as those for X^TX, together with an extra $n - p$ zero eigenvalues. This is easily shown. Let $\mathbf{v_i}$ be an eigen vector of XX^T associated with a non-zero eigen value, and so

$$XX^T\mathbf{v_i} = l_i\mathbf{v_i}$$

Premultiplying by X^T,

$$(X^TX)(X^T\mathbf{v_i}) = l_i(X^T\mathbf{v_i})$$

But

$$X^TX\mathbf{e_i} = \lambda_i e_i$$

and hence $\lambda_i = l_i$ and the eigenvectors are related by $e_i = X^T \mathbf{v_i}$. Thus there is duality between a principal components analysis and a Classical MDS where dissimilarities are given by Euclidean distance. In fact, the coordinates obtained in p' dimensions for the n objects by CMDS are simply the component scores for the n objects on the first p' principal components. Now $\mathbf{e_i}^t \mathbf{e_i} = \mathbf{v_i}^T X X^T \mathbf{v_i} = l_i$. Normalizing $\mathbf{e_i}$, the first p' component scores are given by

$$\begin{aligned} X[l_1^{-1} e_1, l_2^{-1} e_2, ..., l_{p'}^{-1} e_{p'}] &= X[l_1^{-1} X^T \mathbf{v_1}, ..., l_{p'}^{-1} X^T \mathbf{v_{p'}}] \\ &= [l_1^{-\frac{1}{2}} X X^T \mathbf{v_1}, ..., l_{p'}^{-1\frac{1}{2}} X X^T \mathbf{v_{p'}}] \\ &= [l_1^{-\frac{1}{2}} \mathbf{v_1}, ..., l_{p'}^{-1\frac{1}{2}} \mathbf{v_{p'}}] \end{aligned}$$

which are the coordinates obtained from CMDS in p' dimensions.

We have also computed the first two principal components for `iris` data. Figure 3 shows that the principal components coincide with the classical solution of MDS.

Optimal Properties of the Classical Solution and Goodness of Fit

Given a distance matrix D, the object of MDS is to find a configuration $\hat{X}$ in a low-dimensional Euclidean space $\mathbb{R}^p$ whose interpoint distances, $\hat{d}_{rs}^2 = (\hat{\mathbf{x}}_r - \hat{\mathbf{x}}_s)'(\hat{\mathbf{x}}_r - \hat{\mathbf{x}}_s)$ say, closely match D . The "hat" will be used in this section to indicate that the inter-point distances $\hat{D}$ for the configuration $\hat{X}$ are "fitted" to the original distances D. Similarly, let $\hat{B}$ denote the fitted centered inner product matrix.

Now let X be a configuration in $\mathbb{R}^p$ and let $L = (L_1, L_2)$ be a $(p \times p)$ orthogonal matrix where L_1 is $(p \times k)$. Then XL represents a projection of the configuration X onto the subspace of $\mathbb{R}^p$ spanned by the columns of L_1. We can think of $\hat{X} = XL$ as a "fitted" configuration in k dimensions. Since L is orthogonal, the distances between the rows of X are the same as the distances between the rows of XL,

$$\hat{d}_{rs}^2 = \sum_{i=1}^{p} (x_{ri} - x_{si})^2 = \sum_{i=1}^{p} (\mathbf{x_r'} \mathbf{l_{(i)}} - \mathbf{x_s'} \mathbf{l_{(i)}})^2.$$

If we denote teh distances between the row of XL_1 by $\hat{D}$, then

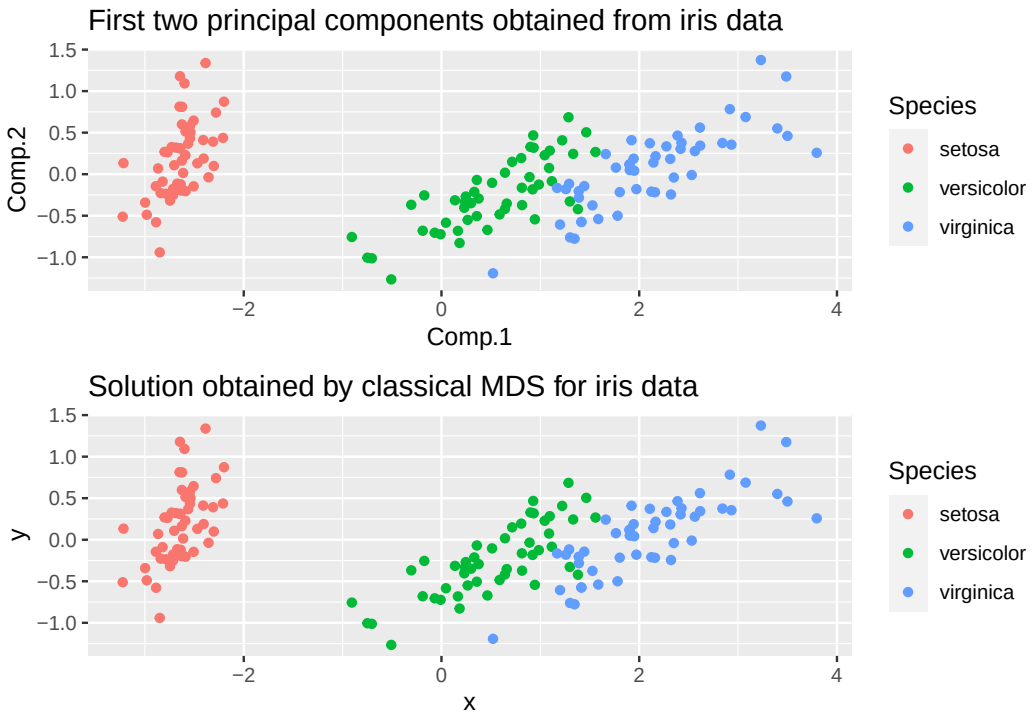

Figure 3: Comparison of PCA and classical MDS solution for iris data

$$\hat{d_{rs}^2} = \sum_{i=1}^{k} (\mathbf{x'_r l_{(i)}} - \mathbf{x'_s l_{(i)}})^2$$

Thus $\hat{d_{rs}} \leq d_{rs}$; that is, projecting a configuration reduces the inter-point distances. Hence, a measure of the discrepancy between the original configuration X and the projected configuration $\hat{X}$ is given by

$$\phi = \sum_{r,s=1}^{n} d_{rs}^2 - \hat{d_{rs}^2}$$

Then the classical solution to the MDS problem in k dimensions has the following optimal property:

Theorem:: *Let* D be a Euclidean distance matrix corresponding to a configuration X in $\mathbb{R}^p$, and fix $k(1 \leq k < p)$. Then amongst all projections XL_1, of X onto k-dimensional subspaces of $\mathbb{R}^p$, the quantity ϕ is minimized when X is projected onto its principal coordinates in k dimensions.

We have already seen that solutions obtained by classical MDS coincide with the principal components. So the solution obtained by classical MDS provides the least measure of discrepancies.

When D is not necessarily Euclidean, it is more convenient to work with the matrix $B = HAH$. If $\hat{X}$ is a fitted configuration with centered inner product matrix $\hat{B}$, then a measure of discrepancy between B and $\hat{B}$ is given by;

$$\psi = \sum_{r,s=1}^{n} (b_{rs} - \hat{b}_{rs})^2 = tr(B - \hat{B})$$

For this measure also, we can prove that the classical solution to the MDS problem is optimal. Let us look at the following theorem.

Theorem: *If* D is a distance matrix (not necessarily Euclidean), the for fixed k, ψ defined earlier is minimized over all configurations $\hat{X}$ in k dimensions when $\hat{X}$ is the classical solution to the MDS problem.

The above two theorems suggest possible *agreement measures* for the "proportion of a distance matrix D explained" by the k-dimensional classical MDS solution. Supposing $\lambda_k > 0$ ($\lambda_1 \geq ...\lambda_n$ denote the eigen values of B), these measures are

$$\alpha_{1,k} = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{n} |\lambda_i|} \times 100\%,$$

and

$$\alpha_{2,k} = \frac{\sum_{i=1}^{k} \lambda_i^2}{\sum_{i=1}^{n} \lambda_i^2} \times 100\%$$

The absolute values of the eigenvalues are considered in the denominator of $\alpha_{1,k}$ because some of the smaller eigen values might be negative.

Example We have calculated the two agreement measures discussed above for the `iris` data.

$$\alpha_{1,2} = 97.7685\%; \quad \alpha_{2,2} = 99.9627\%$$

So we can interpret our obtained results as "*the classical two dimensional classical MDS solution explains a huge proportion of the original dissimilarity matrix*".

In the next section we will discuss about Non-Metric methods of Multidimensional Scaling.

Nonmetric Multidimensional Scaling

Introduction

This section presents the underlying theory of non-metric multidimensional scaling developed in the 1960s.

Suppose there are n objects with dissimilarities $\{\delta_{rs}\}$. The procedure is to find a configuration of n points in a space, which is usually chosen to be Euclidean, so that each object is represented by a point in the space. A configuration is sought so that distances between pairs of points $\{d_{rs}\}$ in the space match "as well as possible" the original dissimilarities $\{\delta_{rs}\}$.

Mathematically, let the objects comprise a set O. Let the dissimilarity, defined on $O \times O$, between objects r and s be δ_{rs} $(r, s \in O)$. Let ϕ be an arbitrary mapping from O onto a set of points X, where X is a subset of the space which is being used to represent the objects. Let the distance between points x_r, x_s in X be given by the real-valued function $d_{x_r x_s}$. Then a disparity, $\hat{d}$, is defined on $O \times O$, which is a measure of how well the distance $d_{\phi(r)\phi(s)}$ matches dissimilarity δ_{rs}. The aim is to find a mapping ϕ, for which $d_{\phi(r)\phi(s)}$ is approximately equal to $\hat{d}_{rs}$ and is usually found by means of some loss function. The points in X together with their associated distances will be referred to as a configuration of points.

The choices are already discussed in previous section, and it is assumed that dissimilarities $\{\delta_{rs}\}$ have been calculated for the set of objects. The set X is often taken as $\mathbb{R}^2$ and d as Euclidean distance, although others are sometimes used, for example $\mathbb{R}^3$, and the Minkowski metric. Once these are chosen, together with the method for calculating disparities, the nonmetric multidimensional scaling problem becomes one of finding an appropriate algorithm for minimizing a loss function.

One choice of d_{rs} can be of the form,

$$d_{rs} = \delta_{rs} + e_{rs}$$

Here the e_{rs} represent errors of measurement plus distortion errors arising because the distances do not exactly correspond to a configuration in $\mathbb{R}^k$.

However, in some situations it is more realistic to hypothesize a less rigid relationship between d_{rs} and δ_{rs}; namely, suppose

$$d_{rs} = f(\delta_{rs} + e_{rs})$$

where f is an unknown monotone increasing function. For this "model", the only information we can use to reconstruct the δ_{rs} is the rank order of the d_{rs}.

In this non-metric approach D is not thought of as a "distance" matrix but as a "dissimilarity" matrix. In fact the non-metric approach is often most appropriate when the data is presented as a similarity matrix. For in this situation the transformation from similarities to distances is somewhat arbitrary and perhaps the strongest statement one should make is that greater similarity implies less dissimilarity.

An algorithm to construct a configuration based on the rank order information has been developed by Shepard and Kruskal. We will now discus the algorithm.

Shepard - Kruskal Algorithm

1. Given a dissimilarity matrix D , order the off-diagonal elements so that

$$d_{r_1 s_1} \leq ... \leq d_{r_m s_m}, \quad m = \frac{1}{2}n(n-1),$$

where $(r_1, s_1), ..., (r_m, s_m)$ denote all pairs of unequal subscripts,$r_i < s_i$. Say that numbers d^*_{rs} are monotonically related to the d_{rs} (and write $d^*_{rs} \overset{mon}{\sim} d_{rs}$) if

$$d_{rs} < d_{uv} \implies d^*_{rs} \leq d^*_{uv} \quad \text{for all} \quad r < s, u < v \qquad ...(*)$$

2. Let $\hat{X}(n \times k)$ be a configuration in $\mathbb{R}^k$ with interpoint distances $\hat{d}_{rs}$. Define the (squared) stress of $\hat{X}$ by

$$S^2(\hat{X}) = \min \frac{\sum_{r<s}(d^*_{rs} - \hat{d}_{rs})^2}{\sum_{r<s} \hat{d}^2_{rs}},$$

where the minimum is taken over d^*_{rs} such that $d^*_{rs} \overset{mon}{\sim} d_{rs}$. The d^*_{rs} which minimizes $S^2(\hat{X})$ represent the least squares least squares monotone regression of $\hat{d}_{rs}$ on d_{rs}. Thus $S^2(\hat{X})$ represents the extent to which the

rank order of the $\hat{d}_{rs}$ disagrees with the rank order of the d_{rs}. If the rank orders match exactly (which is very rare in practice), then $S(\hat{X}) = 0$. The presence of the denominator in $S^2(\hat{X})$ standardizes the stress and makes it invariant under transformations of the sort $\mathbf{y_r} = c\mathbf{x_r}, r = 1, ..., n, c \neq 0$. The stress is also invariant under transformations of the form $\mathbf{y_r} = A\mathbf{x_r}+\mathbf{b}$ when A is orthogonal.

3. For each dimension k, the configuration which has the smallest stress is called the *best fitting configuration in k dimension*. Let

$$S_k = \min_{\hat{X}(n\times k)} S(\hat{X})$$

denote this minimal stress.

4. To choose the correct dimension, calculate $S_1, S_2, ...$, until the value becomes low. Say, for example, S_k is low for $k = k_0$. Since S_k is a decreasing function of k, $k = k_0$ is the "right dimension". A rule of thumb is provided by Kruskal to judge the tolerability of S_k; $S_k \geq 20\%$, poor; $S_k = 10\%$, fair; $S_k \leq 5\%$, good; $S_k = 0$, perfect.

Remarks:

1. The "best configuration" starting from an arbitrary initial configuration can be obtained by using a computer routine developed by Kruskal which utilizes the method of steepest descent to find the local minimum. The initial configuration can be taken as the classical solution. Unfortunately, there is no way of distinguishing in practice between a local minimum and the global minimum.

2. The Shepard-Kruskal solution is invariant under rotation, translation, and uniform expansion or contraction of the best-fining configuration.

3. The Shepard-Kruskal solution is non-metric since it utilizes only the rank orders. However, we still need a sufficiently objective numerical measure of distance to determine the rank order of the d_{rs}.

4. *Similarities* The non-metric method works just as well with similarities as with dissimilarities. One simply changes the direction of the inequalities.

5. *Missing values* The Shepard-Kruskal method is easily adapted to the situation where there are missing values. One simply omits the missing dissimilarities in the ordering and deletes the corresponding terms from the numerator and denominator of $S^2(\hat{X})$. As long as not too many values are missing, the method still seems to work well.

6. *Treatment of ties* The constraint given by $(*)$ is called the primary treatment of lies (PTT). If $d_{rs} = d_{uv}$ then no constraint is made on d^*_{rs} and d^*_{uv}. An alternative constraint, called the *secondary treatment of ties* (STT) is given by
$$d_{rs} \leq d_{uv} \implies d^*_{rs} \leq d^*_{uv},$$
which has the property that $d_{rs} = d_{uv} \implies d^*_{rs} = d^*_{uv}$.

7. *Comparison of methods* The computation is simpler for the classical method than it is for the non-metric method. It is not known how robust the classical method is to monotone transformations of the distance function; however, both methods seem to give similar answers when applied to well-known examples in the field.

Sometimes mostly for calculating the similarity matrix, we would not like to take *Euclidean Distances*. Here is an example.

However, there are cases, particularly in ecological contexts, where a Euclidean Distance is not preferred.

Let's consider an example of species counts for three sites.

Site	Species_1	Species_2	Species_3
A	0	1	1
B	1	0	0
C	0	4	4

If we were to produce the Euclidean distances between each of the sites, it would look something like this:

$$Distance_{AfromB} = \sqrt{(0-1)^2 + (1-0)^2 + (1-0)^2} = 1.732$$

$$Distance_{BfromC} = \sqrt{(1-0)^2 + (0-4)^2 + (0-4)^2} = 5.744$$

$$Distance_{AfromC} = \sqrt{(0-0)^2 + (1-4)^2 + (1-4)^2} = 4.243$$

So, based on these calculated distance metrics, sites A and B are most similar. This conclusion, however, may be counter-intuitive to most ecologists. An ecologist would likely consider sites A and C to be more similar as they contain the same species compositions but differ in the magnitude of individuals. So, an ecologist may require a slightly different metric, such that sites A and C are represented as being more similar.

For this reason, most ecologists use the Bray-Curtis similarity metric, which is defined earlier:

$Distance_{AfromB} = \frac{|0-1|+|1-0|+|1-0|}{|0+1|+|1+0|+|1+0|} = 1$

$Distance_{BfromC} = \frac{|1-0|+|0-4|+|0-4|}{|1+0|+|0+4|+|0+4|} = 1$

$Distance_{AfromC} = \frac{|0-0|+|1-4|+|1-4|}{|0+0|+|1+4|+|1+4|} = 0.6$

Our analysis now shows that sites A and C are most similar, whereas A and C are most dissimilar from B. In general, this is congruent with how an ecologist would view these systems.

An Application to Nonmetric MDS

We have applied the Sheppard Kruskal algorithm over a real life data. The data can found in the following link: https://raw.githubusercontent.com/kpol01/MDS-2024/main/NMDS.csv. The data contain information about various types of grasses found on lands coming from two categories, either Grassland or Marsh.

We have used the `metaMDS` function present in `vegan` package in `R`. The function takes the data frame as input and returns the points obtained by Nonmetric MDS that we need to plot in a two dimensional space. Here we have used the Bray-Kurtis measure as a measure of dissimilarity among the data points. The two dimensional representation of the data as a solution of Nonmetric MDS is shown in Figure 4.

Like CMDS here also we can find a clear separation among the points belonging from the two different categories. This happens because we can expect the points from the same category to be somewhat similar among each other. Thus we can get the advantage of clustering here also.

Shepard Diagram:

In a general nonmetric scaling situation, using the Shepard–Kruskal approach, if we have data $y_1, y_2, ...y_n$ and a model $f_i(\theta)$ with a number of free parameters θ, to choose the parameters in such a way that the rank order of the model approximates the rank order of the data as well as possible we construct a loss function of the form,

$$\sigma(\theta, \hat{y}) = \sum_{i=1}^{n} w_i(\hat{y}_i - f_i(\theta))^2$$

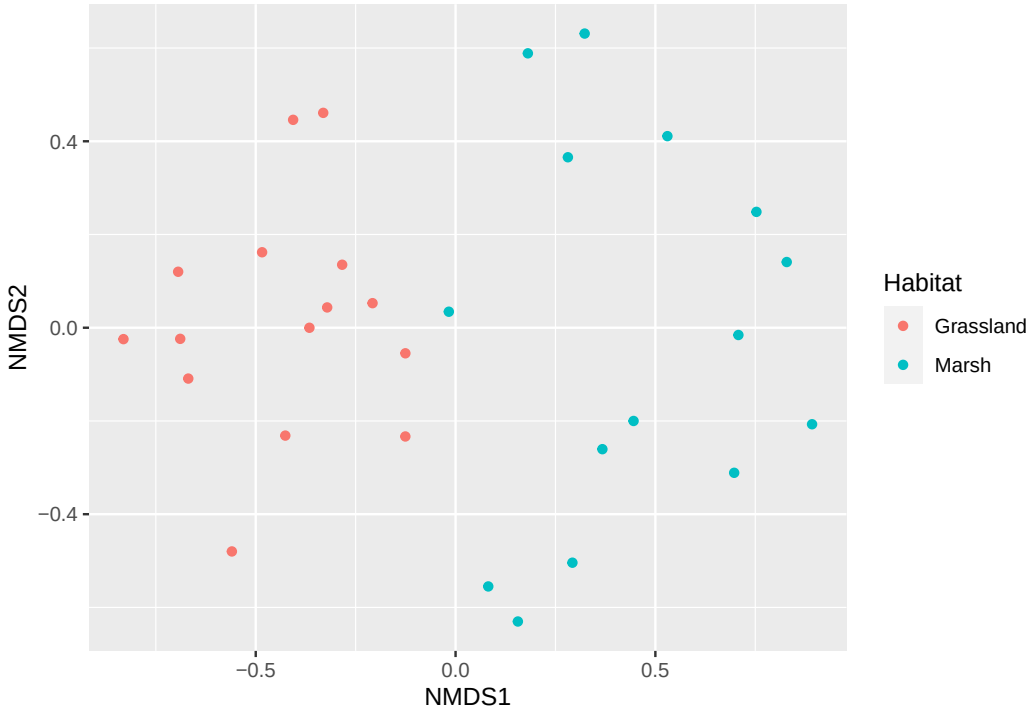

Figure 4: Solution of NMDS obtained from grass data

where w_i are known weights. We then minimize σ over all $\hat{y}$ that are monotone with the data y and over the parameters θ. After we have found the minimum, we plot the data y on the horizontal axis and $\hat{y}$ and the model values f on the vertical axis m to draw the best-fitting monotone step function through the scatter plot. The scatter plot with y and f , and $\hat{y}$ drawn in, is called the *Shepard diagram*. The vertical line of each point from the regression function f gives the residual of this particular point. This gives us an overall picture of the scatter around the regression function including the possibility to detect outliers. In our case, the loss function is,

$$S(\hat{\mathbf{X}}) = \sqrt{\frac{\sum_{r<s}(d^*_{rs} - \hat{d}_{rs})^2}{\sum_{r<s} \hat{d}^2_{rs}}}$$

So plotting d_{rs} on the horizontal axis and d^*_{rs} and $\hat{d}_{rs}$ on the vertical axis, we obtain Shepard Diagram. This shows how the dissimilarities and the approximated distances are related to each other.

We have plotted the Shepard's diagram in Figure 5. In the plot the red line represents a smoothed trend line fitted to the data points. This trend line helps visualize the overall relationship between the observed dissimilarities and the dissimilarities predicted by the NMDS solution.

Linearity : Ideally, the red line should closely follow the diagonal line in the plot, indicating a linear relationship between observed dissimilarities and NMDS-derived dissimilarities. A clear linear relationship suggests that the NMDS solution adequately represents the underlying dissimilarity structure of the data.

Deviation from Diagonal : If the red line deviates significantly from the diagonal line, it suggests that the NMDS solution does not accurately capture the dissimilarity structure of the data. Deviations could indicate issues such as poor model fit or outliers in the data.

Overall Fit : The overall pattern of the red line can provide a visual assessment of the overall fit of the NMDS solution. If the red line closely follows the diagonal line without significant deviations, it indicates a good fit of the NMDS solution to the data.

By examining the red line along with the individual data points in the Shepard plot, researchers can evaluate the goodness-of-fit of the NMDS solution and identify any potential issues or anomalies in the data.

We have plotted the Shepard's diagram in Figure 5. In the plot the red line

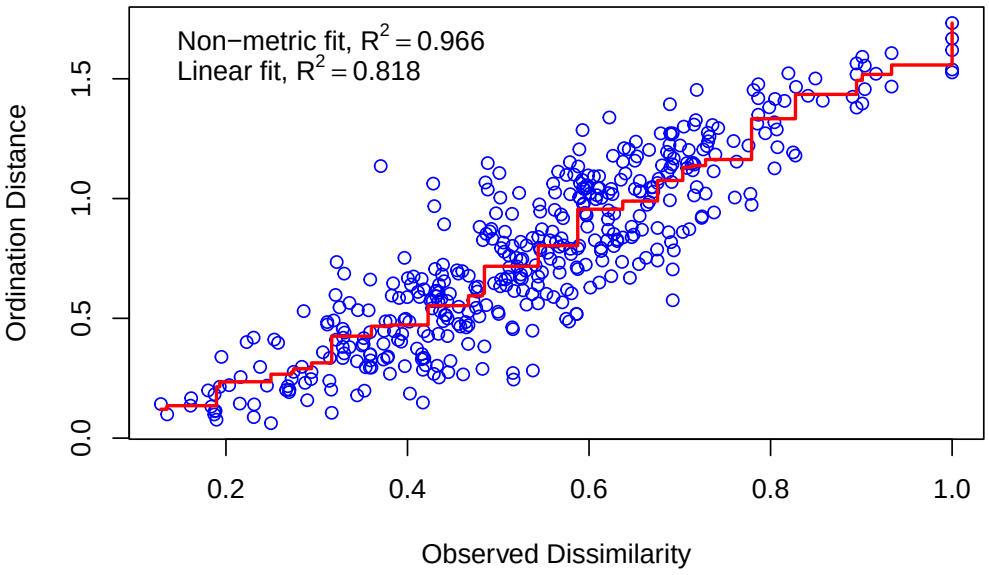

Figure 5: Shepard Diagram for the grass data

represents a smoothed trend line fitted to the data points. This trend line helps visualize the overall relationship between the observed dissimilarities and the dissimilarities predicted by the NMDS solution. Along with the points we are also provided with two goodness of fit measures. A linear fit R^2, which is the coefficient of determination when a linear function of the observed dissimilarities is fitted on the distance between the plotted points. We also get a non-linear fit R^2 which is the coefficient of determination when an optimized non-linear function of the observed dissimilarities is fitted on the distance between the plotted points. From the observed value of both linear and non-linear fit R^2 we can say that we have obtained a very good fit in this case.

Goodness of Fit Measure: Procrustes Rotation

We will now discuss about a goodness of fit measure used to compare two configurations. Let $\mathbf{X}$ be the $(n \times p)$ matrix of the coordinates of n points obtained from $\mathbf{D}$ by one technique. Suppose that V is the $(n \times q)$ matrix of coordinates of another set of points obtained by another technique, or using another measure of distance. Let $q \leq p$. By adding columns of zeros to $\mathbf{Y}$, we may also assume $\mathbf{Y}$ to be $(n \times p)$.

The measure of goodness of fit adopted is obtained by moving the points $\mathbf{y_r}$ relative to the points $\mathbf{x_r}$ until the "residual" sum of squares

$$\sum_{r=1}^{n} (\mathbf{y_r} - \mathbf{x_r})'(\mathbf{y_r} - \mathbf{x_r})$$

is minimal. We can move $\mathbf{y_r}$ relative to $\mathbf{x_r}$ through rotation, reflection and translation, i.e. by

$$\mathbf{A'y_r} + \mathbf{b}, \qquad r = 1, ..., n, \qquad ...(**)$$

where $\mathbf{A'}$ is a $(p \times p)$ orthogonal matrix. Hence, we wish to solve

$$R^2 = \min_{\mathbf{A},\mathbf{b}} \sum_{r=1}^{n} (\mathbf{x_r} - \mathbf{A'y_r} - \mathbf{b})'(\mathbf{x_r} - \mathbf{A'y_r} - \mathbf{b})$$

for $\mathbf{A}$ and $\mathbf{b}$. Note that A and b ar1e found by least squares. Their values are given in the following theorem. But before that we state the *singular value theorem.*

Theorem: (Singular value decomposition theorem) *If* $\mathbf{A}$ *is an* $(nx \times p)$ *matrix of rank* r*, then* $\mathbf{A}$ *can be written as*

$$\mathbf{A} = \mathbf{ULV}'$$

where $\mathbf{U}_{n\times r}$ *and* $\mathbf{V}_{p\times r}$ *are orthogonal matrices* $(\mathbf{U}'\mathbf{U} = \mathbf{V}'\mathbf{V} = \mathbf{I_r})$ *and* $\mathbf{L}$ *is a diagonal matrix with positive elements.*

Theorem: *Let* $X(n \times p)$ *and* $Y(n \times p)$ *be two configurations of* n *points, for convenience centered at the origin, so* $\bar{\mathbf{x}} = \bar{\mathbf{y}} = \mathbf{0}$*. Let* $\mathbf{Z} = \mathbf{Y}'\mathbf{X}$ *and using the singular value decomposition theorem, write*

$$\mathbf{Z} = \mathbf{V\Gamma U}'. \qquad ...(1)$$

where $\mathbf{V}$ *and* $\mathbf{U}$ *are orthogonal* $(p \times p)$ *matrices and* $\mathbf{\Gamma}$ *is a diagonal matrix of non-negative elements. Then the minimizing values of* A *and* $\mathbf{b}$ *for* R^2 *defined earlier are given by*

$$\hat{\mathbf{b}} = \mathbf{0}, \quad \hat{\mathbf{A}} = \mathbf{VU}',$$

and further

$$R^2 = tr(\mathbf{XX}') + tr(\mathbf{YY}') - 2tr(\mathbf{\Gamma}). \qquad ...(2)$$

We have assumed that the column means of $\mathbf{X}$ and $\mathbf{Y}$ are zero. Then the "best" rotation of $\mathbf{Y}$ relative to $\mathbf{X}$ is $\mathbf{Y}\hat{\mathbf{A}}$, where $\hat{\mathbf{A}}$ is as defined in the previous theorem, and $\hat{\mathbf{A}}$ is called the *Procrustes rotation* of $\mathbf{Y}$ relative to $\mathbf{X}$.

Using (1),

$$\mathbf{X}'\mathbf{YY}'\mathbf{X} = \mathbf{Z}'\mathbf{Z} = \mathbf{U\Gamma}^2\mathbf{U}'$$

Thus we can write,

$$R^2 = tr(\mathbf{XX}') + tr(\mathbf{YY}') - 2tr[(\mathbf{X}'\mathbf{YY}'\mathbf{X})^{1/2}]$$

So R^2 is zero if and only if Y can be rotated to X exactly.

If the scales of X and Y are different, then the transformation $(**)$ should be of form,

$$c\mathbf{A}'\mathbf{y_r} + \mathbf{b}, \qquad r = 1, ..., n,$$

where c>0.

The estimate of c is,

$$\hat{c} = tr(\mathbf{\Gamma})/tr(\mathbf{YY}')$$

and the other estimates remain as before. This transformation is called the *Procrustes rotation with scaling* of $\mathbf{Y}$ relative to $\mathbf{X}$ and the new minimum residual sum of squares is given by,

$$R^2 = tr(\mathbf{XX}') + \hat{c}^2 tr(\mathbf{YY}') - 2\hat{c}tr[(\mathbf{X}'\mathbf{YY}'\mathbf{X})^{1/2}].$$

This procedure is not symmetrical with respect to $\mathbf{X}$ and $\mathbf{Y}$. Symmetry can be obtained by selecting scaling so that

$$tr(\mathbf{XX}') = tr(\mathbf{YY}')$$

Some more approaches for MDS:

Metric Least Square Scaling: Metric Least Square Scaling finds a configuration matching d_{rs} toδ_{rs} by minimizing a loss function, S, with possibly a continuous monotonic transformation of dissimilarity, $f(\delta_{rs})$. The configuration $\{x_{ri}\}$ is found in a p-dimensional space but with typically p=2. Sammon (1969), Spaeth and Guthery (1969), Chang and Lee (1973) and Bloxam (1978) are early references. Sammon (1969) suggested the loss function

$$S = \frac{\sum_{r<s} \delta_{rs}^{-1}(d_{rs} - \delta_{rs})^2}{\sum_{r<s} \delta_{rs}}$$

The equations

$$\frac{\partial S}{\partial x_{tk}} = 0 \qquad (t = 1, \cdots, n; k = 1, \cdots, p)$$

have to be solved numerically.

Guttman Approach: Guttman (1968) took a different approach to Kruskal in setting up nonmetric MDS. He defined a loss function called the coefficient of alienation which was basically equivalent to the stress function of Kruskal, but which led to a different algorithm for minimization. His approach will only be described briefly. Let the rank ordered dissimilarities δ_{rs} be placed in a vector δ with elements δ_r $(r = 1, \cdots, N)$. Let the distances $\{\delta_{rs}\}$ from a configuration be placed in a vector $\mathbf{d}$ in order corresponding to $\{\delta_r\}$. Let $\mathbf{E}$ be an NxN permutation matrix which places the elements of $\mathbf{d}$ into ascending order. Disparities are then defined by the rank-image $\mathbf{d}^*$" of $\mathbf{d}$, given by

$$\mathbf{d}^* = \mathbf{E}\mathbf{d}$$

The coefficient of continuity, μ, for the

configuration is given by

$$\mu = \sqrt{\frac{(\sum d_r d_r^*)^2}{\sum d_r^2 \sum d_r^{*2}}}$$

which has the value unity for a perfect fit. In order to find a best fitting configuration, the coefficient of alienation, K, given by

$$K = \sqrt{1 - \mu^2}$$

is minimized using the method of steepest descent. This method is not too much used.

Conclusion

Multidimensional scaling (MDS) is a powerful analytical tool that has revolutionized the way researchers understand and interpret complex data structures. Throughout this exploration, we have delved into the intricacies of MDS, unveiling its theoretical foundations, practical applications, and methodological considerations. As we conclude our journey through the realm of multidimensional scaling, it becomes evident that MDS serves as a versatile and indispensable tool for various disciplines, ranging from psychology and marketing to ecology and geography.

At its core, multidimensional scaling seeks to uncover the underlying geometric structure of data by transforming high-dimensional information into a more

easily interpretable low-dimensional space. Through this process, MDS reveals relationships and patterns that might otherwise remain hidden within the complexity of the original data. Whether applied to perceptual mapping in consumer research, ecological niche modeling in biology, or spatial analysis in geography, MDS offers a means to distill complex information into intuitive visual representations.

One of the key strengths of MDS lies in its ability to preserve the relative distances or dissimilarities between data points. By preserving these relationships, MDS facilitates the exploration of similarities and differences among objects, making it an invaluable tool for clustering, classification, and visualization tasks. Furthermore, MDS can accommodate various types of input data, including similarity/dissimilarity matrices derived from subjective judgments, observational data, or even high-dimensional feature vectors obtained from modern data sources such as genomic sequences or neural activity patterns.

However, the application of MDS is not without its challenges and limitations. Researchers must carefully consider factors such as the choice of distance metric, dimensionality reduction technique, and visualization method, as these decisions can significantly influence the results and interpretation of the analysis. Moreover, the curse of dimensionality poses inherent obstacles when working with high-dimensional data, necessitating careful validation and interpretation of MDS solutions.

Despite these challenges, the benefits of multidimensional scaling are undeniable. By providing a means to uncover latent structures and relationships within complex datasets, MDS enables researchers to gain deeper insights into the underlying mechanisms governing various phenomena. Whether used to understand consumer preferences, species distributions, or spatial configurations, MDS empowers researchers to make informed decisions and formulate hypotheses that drive further inquiry and discovery.

Looking ahead, the future of multidimensional scaling holds promise for continued innovation and refinement. Advances in computational techniques, data visualization, and interdisciplinary collaboration will likely enhance the capabilities and applications of MDS across diverse fields. As we continue to harness the power of multidimensional scaling, we can anticipate new insights, discoveries, and practical solutions to complex problems that shape our understanding of the world around us. In essence, multidimensional scaling stands as a testament to the enduring quest for knowledge and the transformative potential of analytical tools in the pursuit of truth and understanding.

Reference

- Multivariate Analysis; Mardia K. V., Kent J. T., Bibby J. M., United States edition (1995)
- Multidimensional Scaling; Cox T. F., Cox M. A. A., 2nd edition (2000)
- Introduction to Multidimensional Scaling; Schiffman S. S., Reynolds M. L., Young F. W., United Kingdom Edition (1981)
- Modern Multidimensional Scaling; Borg I., Groenen P (1997)
- Multidimensional Scaling: Infinite Metric Measure Spaces; Kassab L (Spring 2019