Chapter 16
Multidimensional Scaling

One major aim of multivariate data analysis is dimension reduction. For data measured in Euclidean coordinates, Factor Analysis and Principal Component Analysis are dominantly used tools. In many applied sciences data is recorded as ranked information. For example, in marketing, one may record "product A is better than product B". High-dimensional observations therefore often have mixed data characteristics and contain relative information (w.r.t. a defined standard) rather than absolute coordinates that would enable us to employ one of the multivariate techniques presented so far.

Multidimensional scaling (MDS) is a method based on proximities between objects, subjects, or stimuli used to produce a spatial representation of these items. Proximities express the similarity or dissimilarity between data objects. It is a dimension reduction technique since the aim is to find a set of points in low dimension (typically 2 dimensions) that reflect the relative configuration of the high-dimensional data objects. The metric MDS is concerned with such a representation in Euclidean coordinates. The desired projections are found via an appropriate spectral decomposition of a distance matrix.

The metric MDS solution may result in projections of data objects that conflict with the ranking of the original observations. The nonmetric MDS solves this problem by iterating between a monotizing algorithmic step and a least squares projection step. The examples presented in this chapter are based on reconstructing a map from a distance matrix and on marketing concerns such as ranking of the outfit of cars.

16.1 The Problem

Multidimensional scaling (MDS) is a mathematical tool that uses proximities between objects, subjects or stimuli to produce a spatial representation of these items. The proximities are defined as any set of numbers that express the amount of similarity or dissimilarity between pairs of objects, subjects or stimuli. In contrast to the techniques considered so far, MDS does not start from the raw multivariate data

W.K. Härdle, L. Simar, *Applied Multivariate Statistical Analysis*,
DOI 10.1007/978-3-642-17229-8_16, © Springer-Verlag Berlin Heidelberg 2012

matrix $\mathcal{X}$, but from a $(n \times n)$ dissimilarity or distance matrix, $\mathcal{D}$, with the elements δ_{ij} and d_{ij} respectively. Hence, the underlying dimensionality of the data under investigation is in general not known.

MDS is a data reduction technique because it is concerned with the problem of finding a set of points in low dimension that represents the "configuration" of data in high dimension. The "configuration" in high dimension is represented by the distance or dissimilarity matrix $\mathcal{D}$.

MDS-techniques are often used to understand how people perceive and evaluate certain signals and information. For instance, political scientists use MDS techniques to understand why political candidates are perceived by voters as being similar or dissimilar. Psychologists use MDS to understand the perceptions and evaluations of speech, colors and personality traits, among other things. Last but not least, in marketing researchers use MDS techniques to shed light on the way consumers evaluate brands and to assess the relationship between product attributes.

In short, the primary purpose of all MDS-techniques is to uncover structural relations or patterns in the data and to represent it in a simple geometrical model or picture. One of the aims is to determine the dimension of the model (the goal is a low-dimensional, easily interpretable model) by finding the d-dimensional space in which there is maximum correspondence between the observed proximities and the distances between points measured on a metric scale.

Multidimensional scaling based on proximities is usually referred to as metric MDS, whereas the more popular nonmetric MDS is used when the proximities are measured on an ordinal scale.

Example 16.1 A good example of how MDS works is given by Dillon and Goldstein (1984) (Page 108). Suppose one is confronted with a map of Germany and asked to measure, with the use of a ruler and the scale of the map, some inter-city distances. Admittedly this is quite an easy exercise. However, let us now reverse the problem: One is given a set of distances, as in Table 16.1, and is asked to recreate the map itself. This is a far more difficult exercise, though it can be solved with a ruler and a compass in two dimensions. MDS is a method for solving this reverse problem in arbitrary dimensions. In Figure 16.2 you can see the graphical representation of the metric MDS solution to Table 16.1 after rotating and reflecting the points representing the cities. Note that the distances given in Table 16.1 are road distances

Table 16.1 Inter-city distances

	Berlin	Dresden	Hamburg	Koblenz	Munich	Rostock
Berlin	0	214	279	610	596	237
Dresden		0	492	533	496	444
Hamburg			0	520	772	140
Koblenz				0	521	687
Munich					0	771
Rostock						0

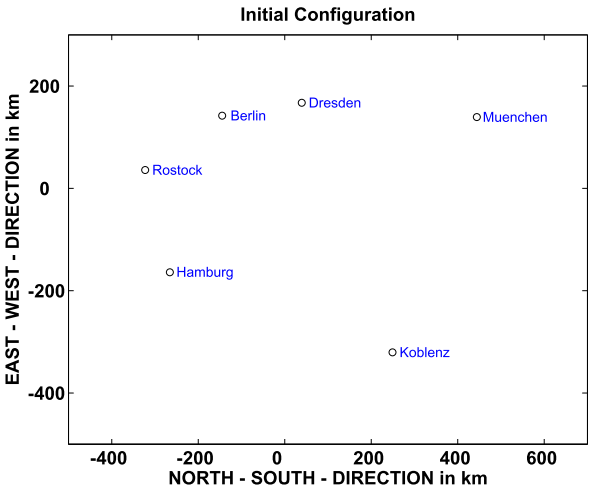


Fig. 16.1 Metric MDS solution for the inter-city road distances `MVAMDScity1`

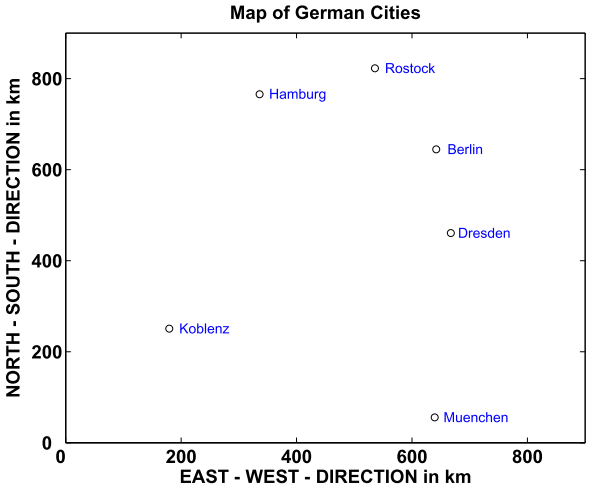


Fig. 16.2 Metric MDS solution for the inter-city road distances after reflection and 90° rotation `MVAMDScity2`

that in general do not correspond to Euclidean distances. In real-life applications, the problems are exceedingly more complex: there are usually errors in the data and the dimensionality is rarely known in advance.

Example 16.2 A further example is given in Table 16.2 where consumers noted their impressions of the dissimilarity of certain cars. The dissimilarities in this table were in fact computed from Table B.7 as Euclidean distances

$$d_{ij} = \sqrt{\sum_{l=1}^{8} (x_{il} - x_{jl})^2}.$$

Table 16.2 Dissimilarities for cars

	Audi 100	BMW 5	Citroen AX	Ferrari	...
Audi 100	0	2.232	3.451	3.689	...
BMW 5	2.232	0	5.513	3.167	...
Citroen AX	3.451	5.513	0	6.202	...
Ferrari	3.689	3.167	6.202	0	...
⋮	⋮	⋮	⋮	⋮	⋱

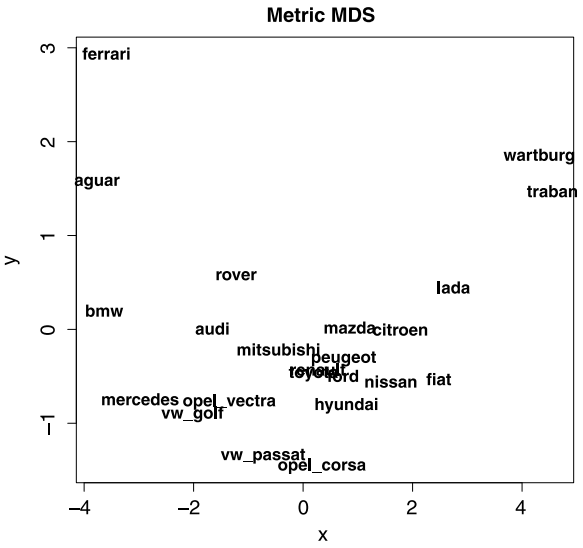


Fig. 16.3 MDS solution on the car data `MVAmdscarm`

MDS produces Figure 16.3 which shows a nonlinear relationship for all the cars in the projection. This enables us to build a nonlinear (quadratic) index with the Wartburg and the Trabant on the left and the Ferrari and the Jaguar on the right. We can construct an order or ranking of the cars based on the subjective impression of the consumers.

What does the ranking describe? The answer is given by Figure 16.4 which shows the correlation between the MDS projection and the variables. Apparently, the first MDS direction is highly correlated with service (−), value (−), design (−), sportiness (−), safety (−) and price (+). We can interpret the first direction as the price direction since a bad mark in price ("high price") obviously corresponds with a good mark, say, in sportiness ("very sportive"). The second MDS direction is highly positively correlated with practicability. We observe from this data an almost orthogonal relationship between price and practicability.

In MDS a map is constructed in Euclidean space that corresponds to given distances. Which solution can we expect? The solution is determined only up to rotation, reflection and shifts. In general, if $P_1, \ldots, P_n$ with coordinates $x_i =$

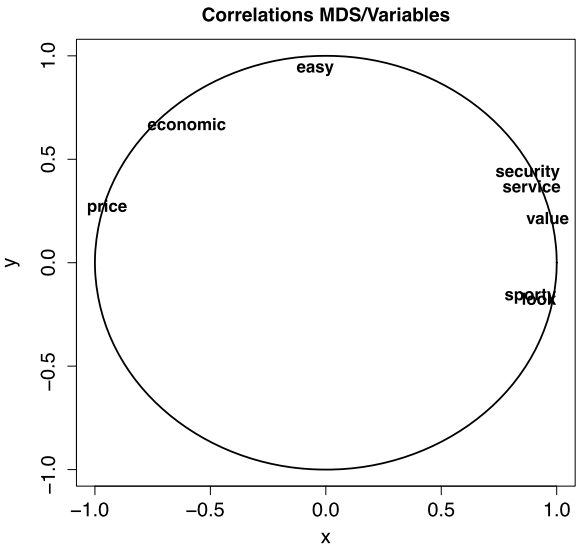
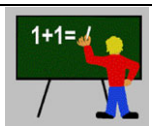


Fig. 16.4 Correlation between the MDS direction and the variables `MVAmdscarm`

$(x_{i1}, \ldots, x_{ip})^\top$ for $i = 1, \ldots, n$ represents a MDS solution in p dimensions, then $y_i = \mathcal{A}x_i + b$ with an orthogonal matrix $\mathcal{A}$ and a shift vector b also represents a MDS solution. A comparison of Figures 16.1 and 16.2 illustrates this fact.

Solution methods that use only the rank order of the distances are termed *nonmetric methods* of MDS. Methods aimed at finding the points P_i directly from a distance matrix like the one in Table 16.2 are called *metric methods*.



Summary

- ↪ MDS is a set of techniques which use distances or dissimilarities to project high-dimensional data into a low-dimensional space essential in understanding respondents perceptions and evaluations for all sorts of items.
- ↪ MDS starts with a $(n \times n)$ proximity matrix $\mathcal{D}$ consisting of dissimilarities $\delta_{i,j}$ or distances d_{ij}.
- ↪ MDS is an explorative technique and focuses on data reduction.
- ↪ The MDS-solution is indeterminate with respect to rotation, reflection and shifts.
- ↪ The MDS-techniques are divided into metric MDS and nonmetric MDS.

16.2 Metric Multidimensional Scaling

Metric MDS begins with a $(n \times n)$ distance matrix $\mathcal{D}$ with elements d_{ij} where i, $j = 1, \ldots, n$. The objective of metric MDS is to find a configuration of points in p-dimensional space from the distances between the points such that the coordinates of the n points along the p dimensions yield a Euclidean distance matrix whose elements are as close as possible to the elements of the given distance matrix $\mathcal{D}$.

The Classical Solution

The classical solution is based on a distance matrix that is computed from a *Euclidean geometry*.

Definition 16.1 A $(n \times n)$ distance matrix $\mathcal{D} = (d_{ij})$ is Euclidean if for some points $x_1, \ldots, x_n \in \mathbb{R}^p$; $d_{ij}^2 = (x_i - x_j)^\top (x_i - x_j)$.

The following result tells us whether a distance matrix is Euclidean or not.

Theorem 16.1 *Define* $\mathcal{A} = (a_{ij})$, $a_{ij} = -\frac{1}{2} d_{ij}^2$ *and* $\mathcal{B} = \mathcal{H}\mathcal{A}\mathcal{H}$ *where* $\mathcal{H}$ *is the centering matrix.* $\mathcal{D}$ *is Euclidean if and only if* $\mathcal{B}$ *is positive semidefinite. If* $\mathcal{D}$ *is the distance matrix of a data matrix* $\mathcal{X}$*, then* $\mathcal{B} = \mathcal{H}\mathcal{X}\mathcal{X}^\top\mathcal{H}$*.* $\mathcal{B}$ *is called the inner product matrix.*

Recovery of Coordinates

The task of MDS is to find the original Euclidean coordinates from a given distance matrix. Let the coordinates of n points in a p dimensional Euclidean space be given by x_i $(i = 1, \ldots, n)$ where $x_i = (x_{i1}, \ldots, x_{ip})^\top$. Call $\mathcal{X} = (x_1, \ldots, x_n)^\top$ the coordinate matrix and assume $\overline{x} = 0$. The Euclidean distance between the i-th and j-th points is given by:

$$d_{ij}^2 = \sum_{k=1}^{p} (x_{ik} - x_{jk})^2. \tag{16.1}$$

The general b_{ij} term of $\mathcal{B}$ is given by:

$$b_{ij} = \sum_{k=1}^{p} x_{ik} x_{jk} = x_i^\top x_j. \tag{16.2}$$

It is possible to derive $\mathcal{B}$ from the known squared distances d_{ij}, and then from $\mathcal{B}$ the unknown coordinates.

$$\begin{aligned} d_{ij}^2 &= x_i^\top x_i + x_j^\top x_j - 2 x_i^\top x_j \\ &= b_{ii} + b_{jj} - 2 b_{ij}. \end{aligned} \tag{16.3}$$

Centering of the coordinate matrix $\mathcal{X}$ implies that $\sum_{i=1}^{n} b_{ij} = 0$. Summing (16.3) over i, over j, and over i and j, we find:

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} d_{ij}^2 &= \frac{1}{n}\sum_{i=1}^{n} b_{ii} + b_{jj} \\ \frac{1}{n}\sum_{j=1}^{n} d_{ij}^2 &= b_{ii} + \frac{1}{n}\sum_{j=1}^{n} b_{jj} \\ \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}^2 &= \frac{2}{n}\sum_{i=1}^{n} b_{ii}. \end{aligned} \tag{16.4}$$

Solving (16.3) and (16.4) gives:

$$b_{ij} = -\frac{1}{2}(d_{ij}^2 - d_{i\bullet}^2 - d_{\bullet j}^2 + d_{\bullet\bullet}^2). \tag{16.5}$$

With $a_{ij} = -\frac{1}{2}d_{ij}^2$, and

$$\begin{aligned} a_{i\bullet} &= \frac{1}{n}\sum_{j=1}^{n} a_{ij} \\ a_{\bullet j} &= \frac{1}{n}\sum_{i=1}^{n} a_{ij} \\ a_{\bullet\bullet} &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} \end{aligned} \tag{16.6}$$

we get:

$$b_{ij} = a_{ij} - a_{i\bullet} - a_{\bullet j} + a_{\bullet\bullet}. \tag{16.7}$$

Define the matrix $\mathcal{A}$ as (a_{ij}), and observe that:

$$\mathcal{B} = \mathcal{H}\mathcal{A}\mathcal{H}. \tag{16.8}$$

The inner product matrix $\mathcal{B}$ can be expressed as:

$$\mathcal{B} = \mathcal{X}\mathcal{X}^\top, \tag{16.9}$$

where $\mathcal{X} = (x_1, \ldots, x_n)^\top$ is the $(n \times p)$ matrix of coordinates. The rank of $\mathcal{B}$ is then

$$\operatorname{rank}(\mathcal{B}) = \operatorname{rank}(\mathcal{X}\mathcal{X}^\top) = \operatorname{rank}(\mathcal{X}) = p. \tag{16.10}$$

As required in Theorem 16.1 the matrix $\mathcal{B}$ is symmetric, positive semidefinite and of rank p, and hence it has p non-negative eigenvalues and $n - p$ zero eigenvalues. $\mathcal{B}$ can now be written as:

$$\mathcal{B} = \Gamma\Lambda\Gamma^\top \tag{16.11}$$

where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_p)$, the diagonal matrix of the eigenvalues of $\mathcal{B}$, and $\Gamma = (\gamma_1, \ldots, \gamma_p)$, the matrix of corresponding eigenvectors. Hence the coordinate matrix $\mathcal{X}$ containing the point configuration in $\mathbb{R}^p$ is given by:

$$\mathcal{X} = \Gamma\Lambda^{\frac{1}{2}}. \tag{16.12}$$

How Many Dimensions?

The number of desired dimensions is small in order to provide practical interpretations, and is given by the rank of $\mathcal{B}$ or the number of nonzero eigenvalues λ_i. If $\mathcal{B}$ is positive semidefinite, then the number of nonzero eigenvalues gives the number of eigenvalues required for representing the distances d_{ij}.

The proportion of variation explained by p dimensions is given by

$$\frac{\sum_{i=1}^{p} \lambda_i}{\sum_{i=1}^{n-1} \lambda_i}. \tag{16.13}$$

It can be used for the choice of p. If $\mathcal{B}$ is not positive semidefinite we can modify (16.13) to

$$\frac{\sum_{i=1}^{p} \lambda_i}{\sum(\text{“positive eigenvalues”})}. \tag{16.14}$$

In practice the eigenvalues λ_i are almost always unequal to zero. To be able to represent the objects in a space with dimensions as small as possible we may modify the distance matrix to:

$$\mathcal{D}^* = d_{ij}^* \tag{16.15}$$

with

$$d_{ij}^* = \begin{cases} 0; & i = j \\ d_{ij} + e \geq 0; & i \neq j \end{cases} \tag{16.16}$$

where e is determined such that the inner product matrix $\mathcal{B}$ becomes positive semidefinite with a small rank.

Similarities

In some situations we do not start with distances but with similarities. The standard transformation (see Chapter 12) from a similarity matrix $\mathcal{C}$ to a distance matrix $\mathcal{D}$ is:

$$d_{ij} = (c_{ii} - 2c_{ij} + c_{jj})^{\frac{1}{2}}. \tag{16.17}$$

Theorem 16.2 *If $\mathcal{C} \leq 0$, then the distance matrix $\mathcal{D}$ defined by (16.17) is Euclidean with centered inner product matrix $\mathcal{B} = \mathcal{H}\mathcal{C}\mathcal{H}$.*

Relation to Factorial Analysis

Suppose that the $(n \times p)$ data matrix $\mathcal{X}$ is centered so that $\mathcal{X}^\top \mathcal{X}$ equals a multiple of the covariance matrix $n\mathcal{S}$. Suppose that the p eigenvalues $\lambda_1, \ldots, \lambda_p$ of $n\mathcal{S}$ are distinct and non zero. Using the duality Theorem 9.4 of factorial analysis we see that $\lambda_1, \ldots, \lambda_p$ are also eigenvalues of $\mathcal{X}\mathcal{X}^\top = \mathcal{B}$ when $\mathcal{D}$ is the Euclidean distance matrix between the rows of $\mathcal{X}$. The k-dimensional solution to the metric MDS problem is thus given by the k first principal components of $\mathcal{X}$.

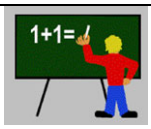
Optimality Properties of the Classical MDS Solution

Let $\mathcal{X}$ be a $(n \times p)$ data matrix with some inter-point distance matrix $\mathcal{D}$. The objective of MDS is thus to find $\mathcal{X}_1$, a representation of $\mathcal{X}$ in a lower dimensional Euclidean space $\mathbb{R}^k$ whose inter-point distance matrix $\mathcal{D}_1$ is not far from $\mathcal{D}$. Let $\mathcal{L} = (\mathcal{L}_1, \mathcal{L}_2)$ be a $(p \times p)$ orthogonal matrix where $\mathcal{L}_1$ is $(p \times k)$. $\mathcal{X}_1 = \mathcal{X}\mathcal{L}_1$ represents a projection of $\mathcal{X}$ on the column space of $\mathcal{L}_1$; in other words, $\mathcal{X}_1$ may be viewed as a fitted configuration of $\mathcal{X}$ in $\mathbb{R}^k$. A measure of discrepancy between $\mathcal{D}$ and $\mathcal{D}_1 = (d_{ij}^{(1)})$ is given by

$$\phi = \sum_{i,j=1}^{n} (d_{ij} - d_{ij}^{(1)})^2. \tag{16.18}$$

Theorem 16.3 *Among all projections $\mathcal{X}\mathcal{L}_1$ of $\mathcal{X}$ onto k-dimensional subspaces of $\mathbb{R}^p$ the quantity ϕ in (16.18) is minimized when $\mathcal{X}$ is projected onto its first k principal factors.*

We see therefore that the metric MDS is identical to principal factor analysis as we have defined it in Chapter 9.



Summary

↪ Metric MDS starts with a distance matrix $\mathcal{D}$.

↪ The aim of metric MDS is to construct a map in Euclidean space that corresponds to the given distances.

↪ A practical algorithm is given as:

1. start with distances d_{ij}
2. define $\mathcal{A} = -\frac{1}{2}d_{ij}^2$
3. put $\mathcal{B} = (a_{ij} - a_{i\bullet} - a_{\bullet j} + a_{\bullet\bullet})$
4. find the eigenvalues $\lambda_1, \ldots, \lambda_p$ and the associated eigenvectors $\gamma_1, \ldots, \gamma_p$ where the eigenvectors are normalized so that $\gamma_i^\top \gamma_i = 1$.
5. Choose an appropriate number of dimensions p (ideally $p = 2$).
6. The coordinates of the n points in the Euclidean space are given by $x_{ij} = \gamma_{ij}\lambda_j^{1/2}$ for $i = 1, \ldots, n$ and $j = 1, \ldots, p$.

↪ Metric MDS is identical to principal components analysis.

16.3 Nonmetric Multidimensional Scaling

The object of nonmetric MDS, as well as of metric MDS, is to find the coordinates of the points in p-dimensional space, so that there is a good agreement between the observed proximities and the inter-point distances. The development of nonmetric MDS was motivated by two main weaknesses in the metric MDS (Fahrmeir and Hamerle, 1984, p. 679):

1. the definition of an explicit functional connection between dissimilarities and distances in order to derive distances out of given dissimilarities, and
2. the restriction to Euclidean geometry in order to determine the object configurations.

The idea of a nonmetric MDS is to demand a less rigid relationship between the dissimilarities and the distances. Suppose that an unknown monotonic increasing function f,

$$d_{ij} = f(\delta_{ij}), \tag{16.19}$$

is used to generate a set of distances d_{ij} as a function of given dissimilarities δ_{ij}. Here f has the property that if $\delta_{ij} < \delta_{rs}$, then $f(\delta_{ij}) < f(\delta_{rs})$. The scaling is based on the rank order of the dissimilarities. Nonmetric MDS is therefore ordinal in character.

The most common approach used to determine the elements d_{ij} and to obtain the coordinates of the objects $x_1, x_2, \ldots, x_n$ given only rank order information is an iterative process commonly referred to as the Shepard-Kruskal algorithm.

Shepard-Kruskal Algorithm

In a first step, called the initial phase, we calculate Euclidean distances $d_{ij}^{(0)}$ from an arbitrarily chosen initial configuration $\mathcal{X}_0$ in dimension p^*, provided that all objects have different coordinates. One might use metric MDS to obtain these initial coordinates. The second step or nonmetric phase determines disparities $\hat{d}_{ij}^{(0)}$ from the distances $d_{ij}^{(0)}$ by constructing a monotone regression relationship between the $d_{ij}^{(0)}$'s and δ_{ij}'s, under the requirement that if $\delta_{ij} < \delta_{rs}$, then $\hat{d}_{ij}^{(0)} \leq \hat{d}_{rs}^{(0)}$. This is called the weak monotonicity requirement. To obtain the disparities $\hat{d}_{ij}^{(0)}$, a useful approximation method is the *pool-adjacent violators* (PAV) algorithm (see Figure 16.5). Let

$$(i_1, j_1) > (i_2, j_2) > \cdots > (i_k, j_k) \tag{16.20}$$

be the rank order of dissimilarities of the $k = n(n-1)/2$ pairs of objects. This corresponds to the points in Figure 16.6. The PAV algorithm is described as follows: "beginning with the lowest ranked value of δ_{ij}, the adjacent $d_{ij}^{(0)}$ values are compared for each δ_{ij} to determine if they are monotonically related to the δ_{ij}'s. Whenever a block of consecutive values of $d_{ij}^{(0)}$ are encountered that violate the required

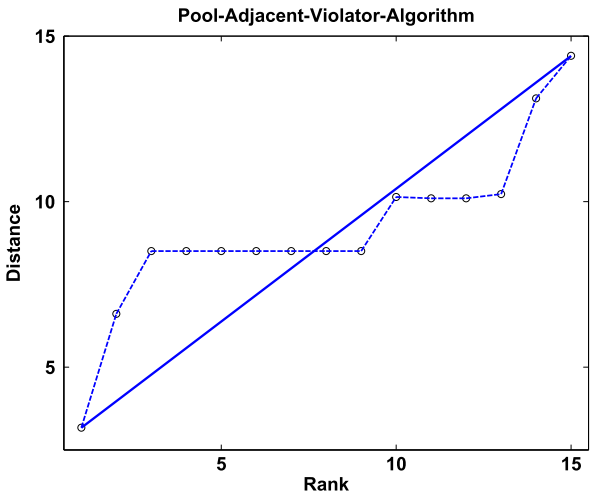


Fig. 16.5 Pool-adjacent violators algorithm `MVAMDSpooladj`

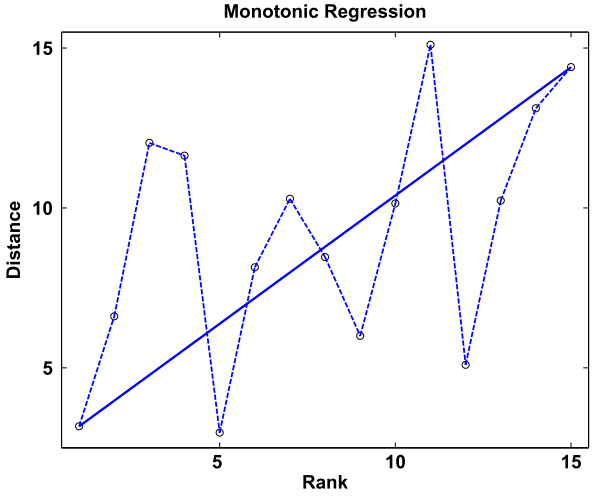


Fig. 16.6 Ranks and distances `MVAMDSnonmstart`

monotonicity property the $d_{ij}^{(0)}$ values are averaged together with the most recent non-violator $d_{ij}^{(0)}$ value to obtain an estimator. Eventually this value is assigned to all points in the particular block".

In a third step, called the metric phase, the spatial configuration of $\mathcal{X}_0$ is altered to obtain $\mathcal{X}_1$. From $\mathcal{X}_1$ the new distances $d_{ij}^{(1)}$ can be obtained which are more closely related to the disparities $\hat{d}_{ij}^{(0)}$ from step two.

Example 16.3 Consider a small example with 4 objects based on the car marks data set, see Table 16.3. Our aim is to find a representation with $p^* = 2$ via MDS. Suppose that we choose as an initial configuration of $\mathcal{X}_0$ the coordinates given in

Table 16.3 Dissimilarities δ_{ij} for car marks

	j	1	2	3	4
i		Mercedes	Jaguar	Ferrari	VW
1	Mercedes	–			
2	Jaguar	3	–		
3	Ferrari	2	1	–	
4	VW	5	4	6	–

Table 16.4 Initial coordinates for MDS

i		x_{i1}	x_{i2}
1	Mercedes	3	2
2	Jaguar	2	7
3	Ferrari	1	3
4	VW	10	4

Table 16.5 Ranks and distances

i, j	d_{ij}	$rank(d_{ij})$	δ_{ij}
1,2	5.1	3	3
1,3	2.2	1	2
1,4	7.3	4	5
2,3	4.1	2	1
2,4	8.5	5	4
3,4	9.1	6	6

Table 16.4. The corresponding distances $d_{ij} = \sqrt{(x_i - x_j)^\top (x_i - x_j)}$ are calculated in Table 16.5.

A plot of the dissimilarities of Table 16.5 against the distance yields Figure 16.8. This relation is not satisfactory since the ranking of the δ_{ij} did not result in a monotone relation of the corresponding distances d_{ij}. We apply therefore the PAV algorithm.

The first violator of monotonicity is the second point $(1, 3)$. Therefore we average the distances d_{13} and d_{23} to obtain the disparities

$$\hat{d}_{13} = \hat{d}_{23} = \frac{d_{13} + d_{23}}{2} = \frac{2.2 + 4.1}{2} = 3.17.$$

Applying the same procedure to $(2, 4)$ and $(1, 4)$ we obtain $\hat{d}_{24} = \hat{d}_{14} = 7.9$. The plot of δ_{ij} versus the disparities $\hat{d}_{ij}$ represents a monotone regression relationship.

In the initial configuration (Figure 16.7), the third point (Ferrari) could be moved so that the distance to object 2 (Jaguar) is reduced. This procedure however also alters the distance between objects 3 and 4. Care should be given when establishing a monotone relation between δ_{ij} and d_{ij}.

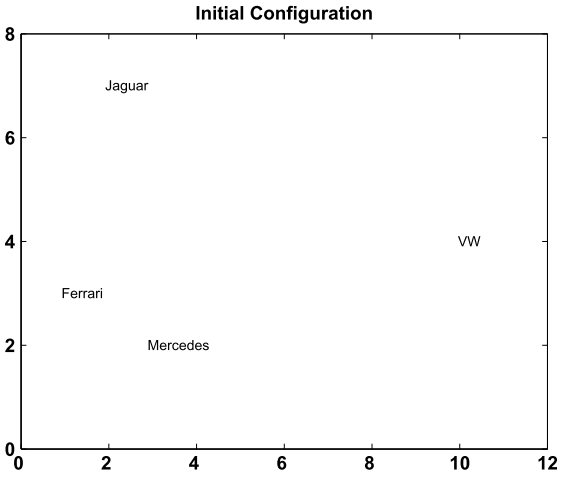


Fig. 16.7 Initial configuration of the MDS of the car data `MVAnmdscar1`

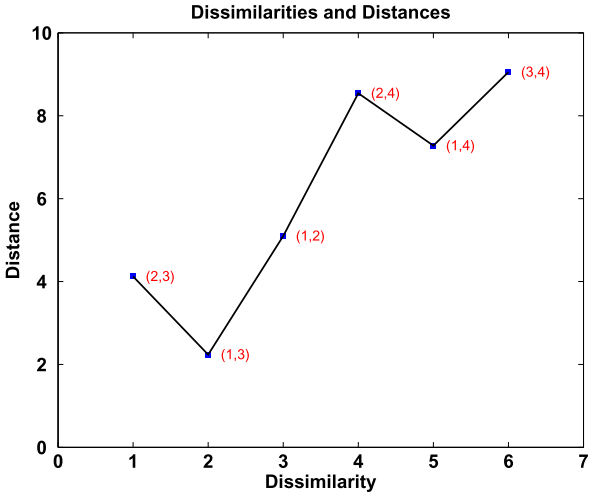


Fig. 16.8 Scatterplot of dissimilarities against distances `MVAnmdscar2`

In order to assess how well the derived configuration fits the given dissimilarities Kruskal suggests a measure called STRESS1 that is given by

$$STRESS1 = \left(\frac{\sum_{i<j} (d_{ij} - \hat{d}_{ij})^2}{\sum_{i<j} d_{ij}^2} \right)^{\frac{1}{2}}. \tag{16.21}$$

An alternative stress measure is given by

$$STRESS2 = \left(\frac{\sum_{i<j} (d_{ij} - \hat{d}_{ij})^2}{\sum_{i<j} (d_{ij} - \overline{d})^2} \right)^{\frac{1}{2}}, \tag{16.22}$$

where $\overline{d}$ denotes the average distance.

Table 16.6 STRESS calculations for car marks example

(i, j)	δ_{ij}	d_{ij}	$\hat{d}_{ij}$	$(d_{ij} - \hat{d}_{ij})^2$	d_{ij}^2	$(d_{ij} - \overline{d})^2$
(2,3)	1	4.1	3.15	0.9	16.8	3.8
(1,3)	2	2.2	3.15	0.9	4.8	14.8
(1,2)	3	5.1	5.1	0	26.0	0.9
(2,4)	4	8.5	7.9	0.4	72.3	6.0
(1,4)	5	7.3	7.9	0.4	53.3	1.6
(3,4)	6	9.1	9.1	0	82.8	9.3
Σ		36.3		2.6	256.0	36.4

Example 16.4 Table 16.6 presents the STRESS calculations for the car example.

The average distance is $\overline{d} = 36.4/6 = 6.1$. The corresponding STRESS measures are:

$$STRESS1 = \sqrt{2.6/256} = 0.1$$
$$STRESS2 = \sqrt{2.6/36.4} = 0.27.$$

The goal is to find a point configuration that balances the effects STRESS and non monotonicity. This is achieved by an iterative procedure. More precisely, one defines a new position of object i relative to object j by

$$x_{il}^{NEW} = x_{il} + \alpha \left(1 - \frac{\hat{d}_{ij}}{d_{ij}}\right)(x_{jl} - x_{il}), \quad l = 1, \ldots, p^*. \tag{16.23}$$

Here α denotes the step width of the iteration.

By (16.23) the configuration of object i is improved relative to object j. In order to obtain an overall improvement relative to all remaining points one uses:

$$x_{il}^{NEW} = x_{il} + \frac{\alpha}{n-1} \sum_{j=1, j\neq i}^{n} \left(1 - \frac{\hat{d}_{ij}}{d_{ij}}\right)(x_{jl} - x_{il}), \quad l = 1, \ldots, p^*. \tag{16.24}$$

The choice of step width α is crucial. Kruskal proposes a starting value of $\alpha = 0.2$. The iteration is continued by a numerical approximation procedure, such as steepest descent or the Newton-Raphson procedure.

In a fourth step, the evaluation phase, the STRESS measure is used to evaluate whether or not its change as a result of the last iteration is sufficiently small that the procedure is terminated. At this stage the optimal fit has been obtained for a given dimension. Hence, the whole procedure needs to be carried out for several dimensions.

Example 16.5 Let us compute the new point configuration for $i = 3$ (Ferrari). The initial coordinates from Table 16.4 are

$$x_{31} = 1 \quad \text{and} \quad x_{32} = 3.$$

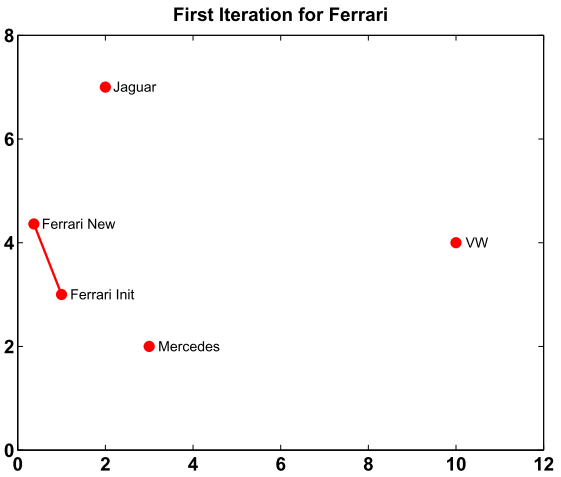


Fig. 16.9 First iteration for Ferrari using Shepard-Kruskal algorithm MVAnmdscar3

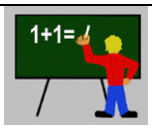
Applying (16.24) yields (for $\alpha = 3$):

$$\begin{aligned} x_{31}^{NEW} &= 1 + \frac{3}{4-1} \sum_{j=1, j\neq 3}^{4} \left(1 - \frac{\hat{d}_{3j}}{d_{3j}}\right)(x_{j1} - 1) \\ &= 1 + \left(1 - \frac{3.15}{2.2}\right)(3-1) + \left(1 - \frac{3.15}{4.1}\right)(2-1) + \left(1 - \frac{9.1}{9.1}\right)(10-1) \\ &= 1 - 0.86 + 0.23 + 0 \\ &= 0.37. \end{aligned}$$

Similarly we obtain $x_{32}^{NEW} = 4.36$.

To find the appropriate number of dimensions, p^*, a plot of the minimum STRESS value as a function of the dimensionality is made. One possible criterion in selecting the appropriate dimensionality is to look for an elbow in the plot. A rule of thumb that can be used to decide if a STRESS value is sufficiently small or not is provided by Kruskal:

$$S > 20\%, \text{ poor}; \quad S = 10\%, \text{ fair}; \quad S < 5\%, \text{ good}; \quad S = 0, \text{ perfect}. \tag{16.25}$$



Summary

↪ Nonmetric MDS is only based on the rank order of dissimilarities.

↪ The object of nonmetric MDS is to create a spatial representation of the objects with low dimensionality.

Summary (continued)

↪ A practical algorithm is given as:

1. Choose an initial configuration.
2. Find d_{ij} from the configuration.
3. Fit $\hat{d}_{ij}$, the disparities, by the PAV algorithm.
4. Find a new configuration $\mathcal{X}_{n+1}$ by using the steepest descent.
5. Go to 2.

16.4 Exercises

Exercise 16.1 Apply the MDS method to the Swiss bank note data. What do you expect to see?

Exercise 16.2 Using (16.6), show that (16.7) can be written in the form (16.2).

Exercise 16.3 Show that

1. $b_{ii} = a_{\bullet\bullet} - 2a_{i\bullet}$; $b_{ij} = a_{ij} - a_{i\bullet} - a_{\bullet j} + a_{\bullet\bullet}$; $i \neq j$
2. $\mathcal{B} = \sum_{i=1}^{p} x_i x_i^\top$
3. $\sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} b_{ii} = \frac{1}{2n \sum_{i,j=1}^{n} d_{ij}^2}$.

Exercise 16.4 Redo a careful analysis of the car marks data based on the following dissimilarity matrix:

	j	1	2	3	4
i		Nissan	Kia	BMW	Audi
1	Nissan	–			
2	Kia	2	–		
3	BMW	4	6	–	
4	Audi	3	5	1	–

Exercise 16.5 Apply the MDS method to the U.S. health data. Is the result in accordance with the geographic location of the U.S. states?

Exercise 16.6 Redo Exercise 16.5 with the U.S. crime data.

Exercise 16.7 Perform the MDS analysis on the Athletic Records data in Appendix B.18. Can you see which countries are "close to each other"?