Data Analysis

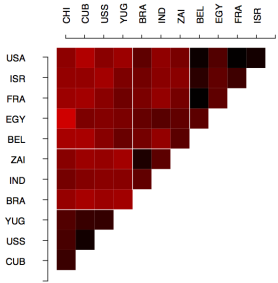
Multidimensional scaling

National Research University Higher School of Economics
Master's Program "Big Data Systems"

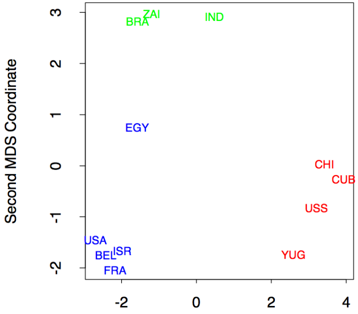
Fall 2019

Goals of multidimensional scaling

- Given pairwise dissimilarities, reconstruct a map that conserves distances:
  - From any dissimilarity (no need to be a metric)
  - Reconstructed map has coordinates $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \dots, x_{im})$ and the natural distance $\|x_i - x_j\|_2$



Reordered Dissimilarity Matrix



First MDS Coordinate

Example

- American electorate's perceptions of thirteen prominent political figures
- Specifically, we have the pairwise dissimilarities between the political figures:
  - With 13 figures, there will be 78 distinct pairs of figures
  - Rank-order pairs of political figures, according to their dissimilarity (from least to most dissimilar).
- For convenience, arrange the rank-ordered dissimilarity values into a square symmetric matrix.

Matrix of Perceptual Dissimilarities among 2004 Political Figures

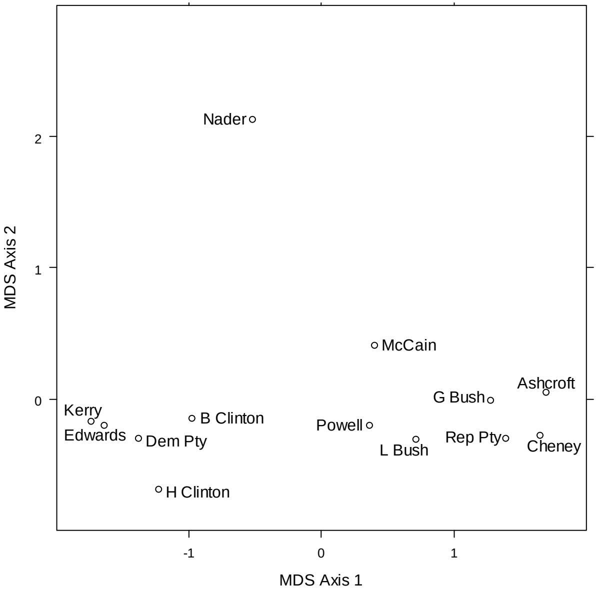
George W Bush	0.0	73.0	62	8.0	68.0	20.0	51.5	41.0	24.0	7	25.5	50	5.0
John Kerry	73.0	0.0	56	78.0	1.0	54.0	15.0	17.0	47.0	77	37.0	2	74.5
Ralph Nader	62.0	56.0	0	72.0	59.0	53.0	60.0	49.0	58.0	70	39.0	57	71.0
Dick Cheney	8.0	78.0	72	0.0	74.5	25.5	65.0	51.5	29.0	12	30.0	66	4.0
John Edwards	68.0	1.0	59	74.5	0.0	44.0	14.0	16.0	46.0	76	38.0	3	69.0
Laura Bush	20.0	54.0	53	25.5	44.0	0.0	42.0	34.0	9.5	23	22.0	45	18.0
Hillary Clinton	51.5	15.0	60	65.0	14.0	42.0	0.0	19.0	32.0	67	40.0	13	55.0
Bill Clinton	41.0	17.0	49	51.5	16.0	34.0	19.0	0.0	31.0	61	36.0	11	48.0
Colin Powell	24.0	47.0	58	29.0	46.0	9.5	32.0	31.0	0.0	28	9.5	35	21.0
John Ashcroft	7.0	77.0	70	12.0	76.0	23.0	67.0	61.0	28.0	0	33.0	63	6.0
John McCain	25.5	37.0	39	30.0	38.0	22.0	40.0	36.0	9.5	33	0.0	43	27.0
Democratic Pty	50.0	2.0	57	66.0	3.0	45.0	13.0	11.0	35.0	63	43.0	0	64.0
Republican Pty	5.0	74.5	71	4.0	69.0	18.0	55.0	48.0	21.0	6	27.0	64	0.0

- Too much information in this matrix to be comprehensible in its “raw” numeric form.
- Instead, try “drawing a picture” of the information in the matrix.

Rules for Drawing the Picture

- Each political figure should be a point in the 2D plane (to enable visualization).
- Find the point locations on that 2D plane so that the rank-ordered of the **distances** (in L_2-sense) between pairs of points corresponds as closely as possible to the rank-ordered **dissimilarities** between pairs of political figures.
- The process of constructing the picture from the dissimilarities is **multidimensional scaling**.

MDS Point Configuration of 2004 Political Figures



Example 2: perception of color in human vision

- 14 colors in the range of 434 μm to 674 μm.
- 31 people rate for each of C_{14}^2 pairs of colors on a five-point scale from 0 (no similarity at all) to 4 (identical).
- Average of 31 ratings for each pair (representing similarity) is then scaled (by $1/4$) and subtracted from 1 to represent dissimilarities.
- The resulting 14×14 dissimilarity matrix is symmetric, and contains zeros in the diagonal.

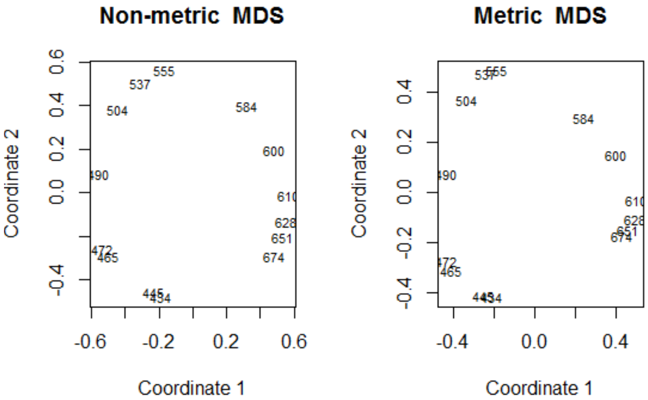
Example 2: perception of color in human vision (2)

	434	445	465	472	490	504	537	555	584	600	610	628	651
445	0.14												
465	0.58	0.50											
472	0.58	0.56	0.19										
490	0.82	0.78	0.53	0.46									
504	0.94	0.91	0.83	0.75	0.39								
537	0.93	0.93	0.90	0.90	0.69	0.38							
555	0.96	0.93	0.92	0.91	0.74	0.55	0.27						
584	0.98	0.98	0.98	0.98	0.93	0.86	0.78	0.67					
600	0.93	0.96	0.99	0.99	0.98	0.92	0.86	0.81	0.42				
610	0.91	0.93	0.98	1.00	0.98	0.98	0.95	0.96	0.63	0.26			
628	0.88	0.89	0.99	0.99	0.99	0.98	0.98	0.97	0.73	0.50	0.24		
651	0.87	0.87	0.95	0.98	0.98	0.98	0.98	0.98	0.80	0.59	0.38	0.15	
674	0.84	0.86	0.97	0.96	1.00	0.99	1.00	0.98	0.77	0.72	0.45	0.32	0.24

- MDS seeks a 2D representation of the table.

Example 2: perception of color in human vision (3)

- MDS reproduces the well-known two-dimensional color circle.



Utility of MDS for research

- Reducing dimensionality.
- Modeling perceptions of survey respondents or experimental subjects.
- Flexible with respect to input data.
- Useful measurement tool.
- Graphical output.

The map analogy:

- A familiar task:
  - Starting with a map (a geometric model), obtain the distances between locations (numeric data)
- MDS “reverses” the preceding task:
  - Start with distances (numeric data) and produce a map (geometric model).

Similarities and dissimilarities

- Given a $n \times n$ symmetric matrix $\mathbf{\Delta}$ of proximities between n objects
- The proximity between the object represented by the i-th row and the object represented by the j-th column is shown by the cell entry δ_{ij}.
- Greater proximity between objects i and j corresponds to smaller values of δ_{ij} and vice versa.
- Therefore, the proximities are often called "dissimilarities".
- The latter may be confusing!

Distances

- In mathematics, a distance function (that gives a distance between two objects x and y) is also called metric, satisfying:

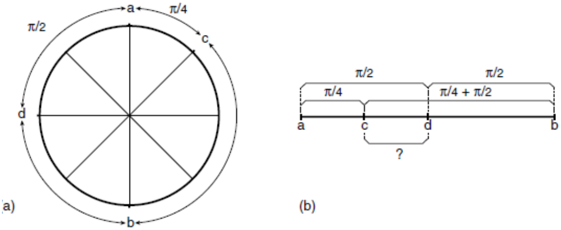
  1. $d(x, y) \geq 0$.
  2. $d(x, y) = 0$ <=> $x = y$.
  3. $d(x, y) = d(y, x)$.
  4. $d(x, z) \leq d(x, y) + d(y, z)$.

- Given a set of dissimilarities, $\boldsymbol{\Delta}$, one wonders whether these values are distances and, moreover, ...
- ...whether they can be interpreted as Euclidean distances.
- The Euclidean distance between $\boldsymbol{x}_i, \boldsymbol{x}_j \in \mathbb{R}^p$ reads:

$$d_{ij} = \left\| \boldsymbol{x}_i - \boldsymbol{x}_j \right\|_2 = \sqrt{\sum_{k=1}^{p} (x_{ik} - x_{jk})^2}.$$

Non-Euclidean distances

- Radian distance (=arc length) function on a circle is a metric.
- Cannot be embedded in $\mathbb{R}^1$ (in other words, one cannot find $x_1, \ldots, x_4 \in \mathbb{R}^1$ to match the distance):



Point	a	b	c	d
a	0.0000	3.1416	0.7854	1.5708
b	3.1416	0.0000	2.3562	1.5708
c	0.7854	2.3562	0.0000	2.3562
d	1.5708	1.5708	2.3562	0.0000

MDS: formal objective

- Given a $n \times n$ dissimilarity matrix $\boldsymbol{\Delta} = \{\delta_{ij}\}$, the MDS seeks for points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^p$ (called a configuration) so that

$$\delta_{ij} \approx \left\| \boldsymbol{x}_i - \boldsymbol{x}_j \right\|_2$$

  for all i, j.
- p should be small ($p = 2, 3$) to enable visualization.
- Sometimes, perhaps, for large p, there exists a configuration

$$\delta_{ij} = \left\| \boldsymbol{x}_i - \boldsymbol{x}_j \right\|_2$$

  for all i,j. Such dissimilarity matrix $\boldsymbol{\Delta}$ is called Euclidean.
- Sometimes, for any p,

$$\delta_{ij} \neq \left\| \boldsymbol{x}_i - \boldsymbol{x}_j \right\|_2$$

  for some i and j. Such dissimilarity matrix $\boldsymbol{\Delta}$ is called non-Euclidean.

MDS: formal objective (2)

- More generally, the MDS uses the information contained in $\boldsymbol{\Delta}$ to find $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^p$ such that the L_2-distances in $\mathbb{R}^p$ are functionally related to the pairwise dissimilarities, i.e., for all pairs $i \neq j$:
$$d_{ij} = f(\delta_{ij}).$$
- The function $f(\cdot)$ is determined by the type of the MDS (metric or non-metric).

Classical multidimensional scaling: theory

- Given an Euclidean $n \times n$ dissimilarity $\boldsymbol{\Delta}$, find n q-dimensional ($q \leq n$) vectors (configuration)

$$\mathbf{X} = \left\{ \underset{\downarrow}{\boldsymbol{x}_1}, \ldots, \underset{\downarrow}{\boldsymbol{x}_n} \right\} \tag{1}$$

such that

$$\delta_{ij} = \left\| \boldsymbol{x}_i - \boldsymbol{x}_j \right\|_2.$$

- Multiple solutions arise as for $\boldsymbol{x}_i^* = \boldsymbol{x}_i + \boldsymbol{c}$:

$$\left\| \boldsymbol{x}_i^* - \boldsymbol{x}_j^* \right\|_2 = \left\| \boldsymbol{x}_i - \boldsymbol{x}_j \right\|_2 = \delta_{ij}$$

- Additional constraint:

$$\sum_{i=1}^{n} \boldsymbol{x}_i = \boldsymbol{0}. \tag{2}$$

Classical multidimensional scaling: theory (2)

- Using (1), build the $n \times n$ Gram matrix $\mathbf{B} = \mathbf{X}^T\mathbf{X}$ of inner products.
- Denote the entries as $\mathbf{B} = \{b_{ij}\}$.
- As $\delta_{ij}^2 = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2 = \|\boldsymbol{x}_i\|^2 + \|\boldsymbol{x}_j\|^2 - 2\boldsymbol{x}_i\boldsymbol{x}_j$, we have

$$\delta_{ij}^2 = b_{ii} + b_{jj} - 2b_{ij}. \tag{3}$$

- Recall, δ_{ij} are given and b_{ij} are sought for.
- Constraint (2) leads to:

$$\sum_{i=1}^{n} b_{ij} = \sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_j = \sum_{i=1}^{n}\sum_{k=1}^{q} x_{ik}x_{jk} = \sum_{k=1}^{q} x_{jk} \sum_{i=1}^{n} x_{ik} = 0,$$

for $j = 1, \dots, n$.
- That is, the sum over each row or column of $\mathbf{B}$ is zero.
- Denoting $T = \text{trace}(\mathbf{B}) = \sum_{i=1}^{n} b_{ii}$ and using (3) we have

$$\sum_{i=1}^{n} \delta_{ij}^2 = T + nb_{jj}, \quad \sum_{j=1}^{n} \delta_{ij}^2 = T + nb_{ii}, \quad \sum_{j=1}^{n}\sum_{i=1}^{n} \delta_{ij}^2 = 2nT. \tag{4}$$

Classical multidimensional scaling: theory (3)

- Combining (3) and (4), we get a unique solution for b_{ij}:

$$b_{ij} = 1/2\left(\delta_{ij}^2 - \delta_{\bullet j}^2 - \delta_{i\bullet}^2 + \delta_{\bullet\bullet}^2\right) \tag{5}$$

  where $\delta_{\bullet j}^2$,$\delta_{i\bullet}^2$ are the averages over i and j, respectively, and $\delta_{\bullet\bullet}^2$ is the average over both i and j.
- The configuration $\mathbf{X}$ is given by eigen-decomposition of $\mathbf{B} = \mathbf{X}^T\mathbf{X}$
- Matrix $\mathbf{B}$ is known, see (5), and symmetrical => can be diagonalized in the basis comprised of its eigenvectors:

$$\mathbf{V}^T\mathbf{B}\mathbf{V} = \mathbf{\Lambda} \tag{6}$$

  where $\mathbf{\Lambda} = \text{diag}\{\lambda_1, ..., \lambda_n\}$ is diagonal and all λ_i being non-negative.
- The rows of matrix $\mathbf{V}$ are the eigenvectors of $\mathbf{B}$.
- Equation (6) is equivalent to:

$$\mathbf{B} = \mathbf{X}^T\mathbf{X} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T = \mathbf{V}\mathbf{\Lambda}^{1/2}\mathbf{\Lambda}^{1/2}\mathbf{V}^T \tag{7}$$

Classical multidimensional scaling: theory (4)

- From (7), we get the configuration $\mathbf{X}$:

$$\mathbf{X} = \mathbf{\Lambda}^{1/2}\mathbf{V}^T \tag{8}$$

- The above procedure is equivalent to the PCA on centered $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$.
- The first PC direction (column) of $\mathbf{X}$ has the largest variation.
- If we wish to reduce the dimension to $p < n$, then the first p rows of $\mathbf{X}$, $\mathbf{X}_{(p)}$, conserve the dissimilarities δ_{ij} best of all other linear dimension reductions of $\mathbf{X}$

$$\mathbf{X}_{(p)} = \mathbf{\Lambda}_p^{1/2}\mathbf{V}_p^T$$

where $\mathbf{\Lambda}_p^{1/2}$ is the first $p \times p$ submatrix of $\mathbf{\Lambda}$ and $\mathbf{V}_p$ is the first p columns of $\mathbf{V}$.
- The dissimilarities are conserved just approximately:

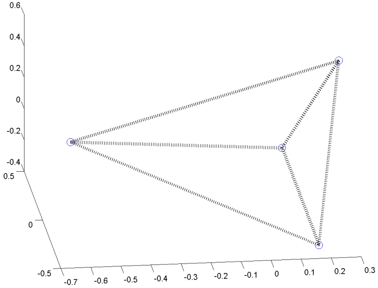
$$\delta_{ij}^2 = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2 = \left\|\boldsymbol{x}_i^{(p)} - \boldsymbol{x}_j^{(p)}\right\|^2 + \underbrace{\left\|\boldsymbol{x}_i^{(*)} - \boldsymbol{x}_j^{(*)}\right\|^2}_{neglect}$$

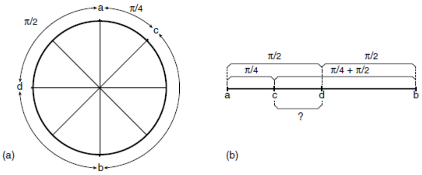
Classical multidimensional scaling: remarks

- Classical MDS gives configurations $\mathbf{X}_{(p)}$ in $\mathbb{R}^p$ for any dimension $1 \leq p \leq n$.
- Configuration is centered.
- Coordinates are given by the principal scores, ordered from largest-to-smallest variation.
- Dimension reduction from $\mathbf{X}$ to $\mathbf{X}_{(p)}$, $(p < n)$, is the same as for the PCA (cutting some PC scores out).
- Leads to exact solution if the dissimilarity is based on the Euclidean distances.
- Can also be used for non-Euclidean distances, in fact, for any dissimilarities.

Classical MDS: examples

- Consider two working examples:
  - the Euclidean geometry (tetrahedron, unit edge length)
  - the circular geometry





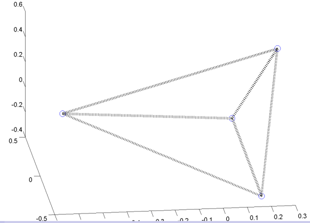
(a) (b)

Classical MDS examples: tetrahedron

- Pairwise dissimilarity matrix for tetrahedron (comprised of unit distances between the points)

$$\begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

- Yields the Gram matrix $\mathbf{B}_{(4\times4)}$ with eigenvalues $(0.5, 0.5, 0.5, 0)$.
- Using dimension $p = 3$, we perfectly restore the tetrahedron.



Classical MDS examples: circular distances

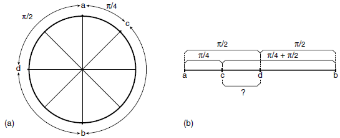
- Pairwise dissimilarity matrix $\boldsymbol{\Delta}$:

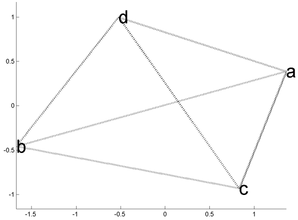
Point	a	b	c	d
a	0.0000	3.1416	0.7854	1.5708
b	3.1416	0.0000	2.3562	1.5708
c	0.7854	2.3562	0.0000	2.3562
d	1.5708	1.5708	2.3562	0.0000

- Yields the Gram matrix $\mathbf{B}_{(4\times4)}$ with eigenvalues $(5.6117, -1.2039, -0.0000, 2.2234)$.
- In retrieving the coordinate matrix $\mathbf{X}$, we cannot take a square root of $\boldsymbol{\Lambda}$ since it yields complex numbers.
- Remedy: Keep only positive eigenvalues and corresponding coordinates. In this case, take coordinates 1 and 4.
- This is the price we pay to represent the non-Euclidean geometry by the Euclidean one.

Classical MDS examples: circular distances (2)

- Using $p = 2$ (cannot use p > 2), configuration $\mathbf{X}_{(2)}$ is:





- Compare the original dissimilarity matrix $\mathbf{\Delta}$ and approximate distance matrix:

$$\begin{pmatrix} 0 & 3.1416 & 0.7854 & 1.5708 \\ 3.1416 & 0 & 2.3562 & 1.5708 \\ 0.7854 & 2.3562 & 0 & 2.3562 \\ 1.5708 & 1.5708 & 2.3562 & 0 \end{pmatrix}, \quad \hat{D} = \begin{pmatrix} 0 & 3.1489 & 1.4218 & 1.9784 \\ 3.1489 & 0 & 2.5482 & 1.8557 \\ 1.4218 & 2.5482 & 0 & 2.3563 \\ 1.9784 & 1.8557 & 2.3563 & 0 \end{pmatrix}$$

Classical MDS examples: airline distances

	Beijing	Cape Town	Hong Kong	Honolulu	London	Melbourne
Cape Town	12947					
Hong Kong	1972	11867				
Honolulu	8171	18562	8945			
London	8160	9635	9646	11653		
Melbourne	9093	10338	7392	8862	16902	
Mexico	12478	13703	14155	6098	8947	13557
Montreal	10490	12744	12462	7915	5240	16730
Moscow	5809	10101	7158	11342	2506	14418
New Delhi	3788	9284	3770	11930	6724	10192
New York	11012	12551	12984	7996	5586	16671
Paris	8236	9307	9650	11988	341	16793
Rio de Janeiro	17325	6075	17710	13343	9254	13227
Rome	8144	8417	9300	12936	1434	15987
San Francisco	9524	16487	11121	3857	8640	12644
Singapore	4465	9671	2575	10824	10860	6050
Stockholm	6725	10334	8243	11059	1436	15593
Tokyo	2104	14737	2893	6208	9585	8159

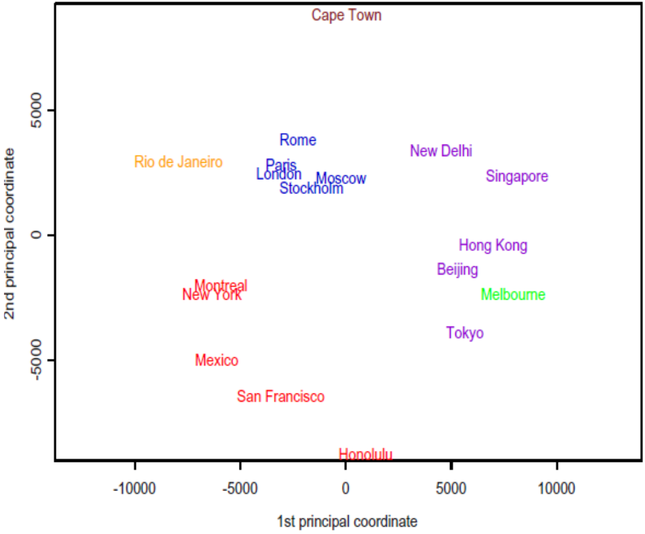
	Mexico	Montreal	Moscow	New Delhi	New York	Paris
Montreal	3728					
Moscow	10740	7077				
New Delhi	14679	11286	4349			
New York	3362	533	7530	11779		
Paris	9213	5522	2492	6601	5851	

Classical MDS examples: airline distances (2)

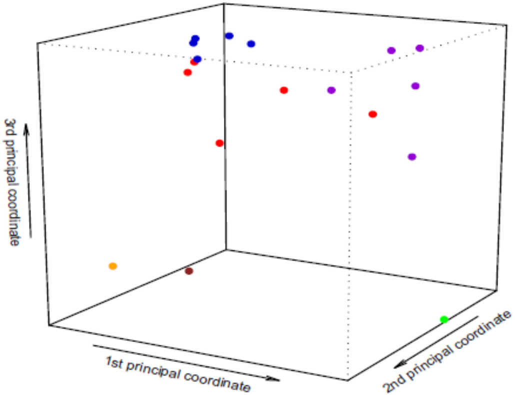
- Airline distances are non-Euclidean
- Take the first 3 largest eigenvalues

	Eigenvalues	Eigenvectors		
1	471582511	0.245	−0.072	0.183
2	316824787	0.003	0.502	-0.347
3	253943687	0.323	−0.017	0.103
4	−98466163	0.044	−0.487	-0.080
5	−74912121	−0.145	0.144	0.205
6	−47505097	0.366	−0.128	−0.569
7	31736348	−0.281	−0.275	−0.174
8	−7508328	−0.272	−0.115	0.094
9	4338497	−0.010	0.134	0.202
10	1747583	0.209	0.195	0.110
11	−1498641	−0.292	−0.117	0.061
12	145113	−0.141	0.163	0.196
13	−102966	−0.364	0.172	−0.473
14	60477	−0.104	0.220	0.163
15	−6334	−0.140	−0.356	−0.009
16	−1362	0.375	0.139	−0.054
17	100	−0.074	0.112	0.215
18	0	0.260	−0.214	0.173

Classical MDS examples: airline distances (3)



Classical MDS examples: airline distances (4)



Classical MDS examples: driving distances

- Driving distances between ten American cities:

0	0.587	1.212	0.701	1.936	0.604	0.748	2.139	2.182	0.543	ATLANTA
0.587	0	0.920	0.940	1.745	1.188	0.713	1.858	1.737	0.597	CHICAGO
1.212	0.920	0	0.879	0.831	1.726	1.631	0.949	1.021	1.494	DENVER
0.701	0.940	0.879	0	1.374	0.968	1.420	1.645	1.891	1.220	HOUSTON
1.936	1.745	0.831	1.374	0	2.339	2.451	0.347	0.959	2.300	LOS ANGELES
0.604	1.188	1.726	0.968	2.339	0	1.092	2.594	2.734	0.923	MIAMI
0.748	0.713	1.631	1.420	2.451	1.092	0	2.571	2.408	0.205	NEW YORK
2.139	1.858	0.949	1.645	0.347	2.594	2.571	0	0.678	2.442	SAN FRANCISCO
2.182	1.737	1.021	1.891	0.959	2.734	2.408	0.678	0	2.329	SEATTLE
0.543	0.597	1.494	1.229	2.300	0.923	0.205	2.442	2.329	0	WASHINGTON DC

Classical MDS examples: driving distances (2)

- Matrix $\mathbf{B}$

```
 0.537  0.228 -0.348  0.199 -0.808  0.895  0.697 -1.005 -1.050  0.656
 0.228  0.263 -0.174 -0.134 -0.594  0.234  0.585 -0.581 -0.315  0.488
-0.348 -0.174  0.236 -0.092  0.570 -0.563 -0.504  0.681  0.658 -0.463
 0.199 -0.134 -0.092  0.352  0.029  0.516 -0.124 -0.163 -0.550 -0.033
-0.808 -0.594  0.570  0.029  1.594 -1.130 -1.499  1.751  1.399 -1.313
 0.895  0.234 -0.563  0.516 -1.130  1.617  0.920 -1.542 -1.867  0.918
 0.697  0.585 -0.504 -0.124 -1.499  0.920  1.416 -1.583 -1.130  1.222
-1.005 -0.581  0.681 -0.163  1.751 -1.542 -1.583  2.028  1.846 -1.432
-1.050 -0.315  0.658 -0.550  1.399 -1.867 -1.130  1.846  2.124 -1.115
 0.656  0.488 -0.463 -0.033 -1.313  0.918  1.222 -1.432 -1.115  1.071
```

Classical MDS examples: driving distances (3)

First two eigenvectors of double-centered data matrix, $\mathbf{\Delta}^*$:

-0.23217	-0.11011
-0.12340	0.26253
0.15554	0.01929
-0.05216	-0.44079
0.38889	-0.30037
-0.36618	-0.44802
-0.34640	0.39964
0.45892	-0.08658
0.43346	0.44649
-0.31645	0.25843

First two eigenvalues of double-centered data matrix, $\mathbf{\Delta}^*$:

9.58217	1.68664

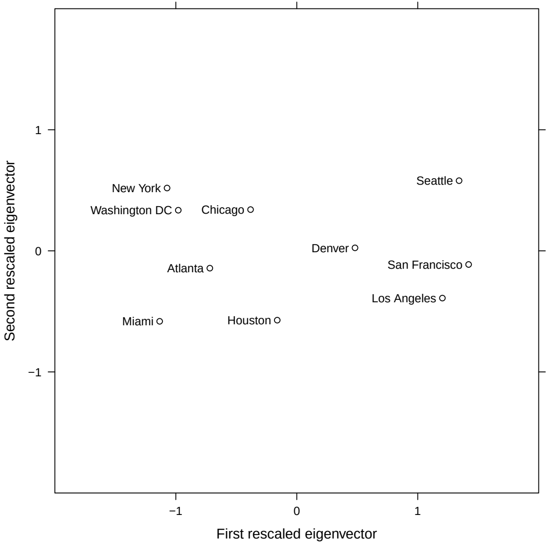
Classical MDS examples: driving distances (4)

- The eigenvectors are multiplied by the square roots of the corresponding eigenvalues to produce the matrix of two-dimensional point coordinates, see (8):

-0.71867	-0.14300	Atlanta
-0.38197	0.34095	Chicago
0.48149	0.02505	Denver
-0.16147	-0.57246	Houston
1.20382	-0.39009	Los Angeles
-1.13352	-0.58185	Miami
-1.07228	0.51901	New York
1.42058	-0.11244	San Francisco
1.34179	0.57986	Seattle
-0.97958	0.33562	Washington D.C.

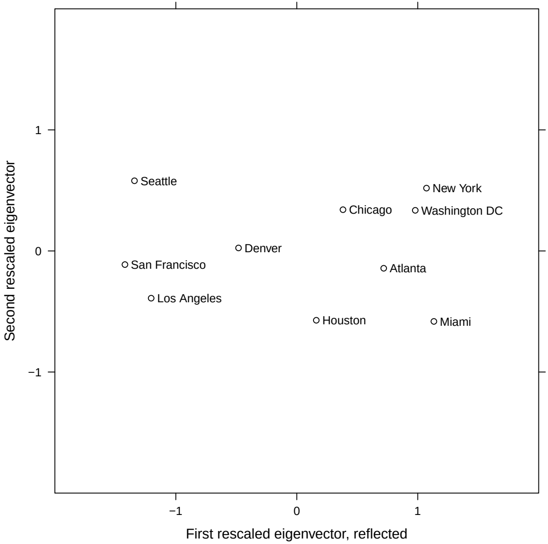
- Point coordinates can be plotted in two-dimensional space.

Classical MDS examples: driving distances (4)



Classical MDS examples: driving distances (5)

- Reflecting the first axis:



Metric MDS

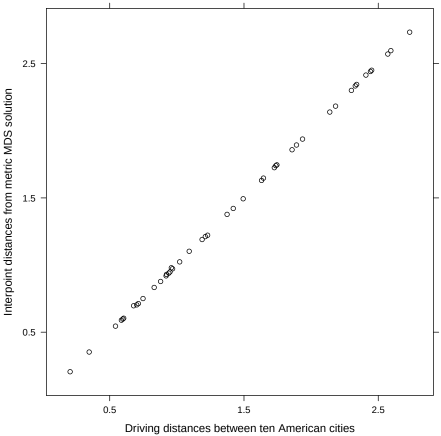
- Metric multidimensional scaling requires that distances are related to dissimilarities by a linear function:

$$d_{ij} = a + b\delta_{ij} + e_{ij}$$

  where a and b are coefficients to be estimated, and e_{ij} is an error term associated with objects i and j.

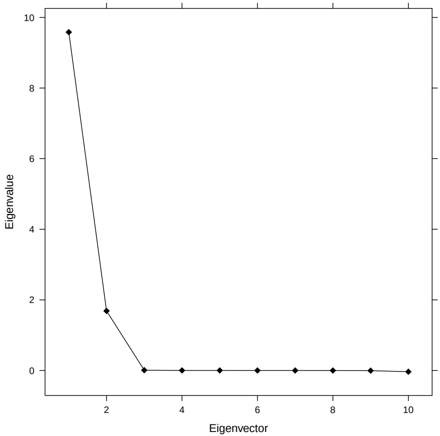
Classical multidimensional scaling examples: driving distances (6)

- Graph of scaled interpoint distances versus input dissimilarities data (i.e., driving distances):



Classical multidimensional scaling examples: driving distances (6)

- Graph of eigenvalues versus order of extraction in metric MDS of intercity driving distances:



Goodness of Fit: measure for metric MDS

- Eigenvalues measure variance associated with each dimension of the MDS solution.
- Sum of first p eigenvalues relative to sum of all eigenvalues:

$$\text{Fit} = \frac{\sum_{i=1}^{p} \lambda_i^2}{\sum_{i=1}^{q} \lambda_i^2}$$

- The first two eigenvalues are 9.58 and 1.69, and the sum of the eigenvalues is 11.32:

$$\text{Fit} = \frac{9.58 + 1.69}{11.32} = 0.996$$

.

Conceptual distances

- If MDS works for physical distances, then it may also work for data that can be interpreted as “conceptual distances”.
- One important type of conceptual distance data:
  - Each of n objects has scores on each of K variables.
  - Each object’s vector of scores is called its profile.
  - For objects i and j, each of which have scores on variables $x_1, x_2, ..., x_K$, the profile dissimilarity is:

$$\delta_{ij} = \sqrt{\sum_{s=1}^{K} (x_{is} - x_{js})^2}$$

  - δ_{ij} is the distance between i and j in the K-dimensional space.

Conceptual distances: example

- Socioeconomic characteristics of ten American cities

	Climate, Terrain	Housing	Environ., Health	Crime	Transport-ation	Education	The Arts	Recreation	Economics
Atlanta	0.185	-1.338	-0.451	-0.609	0.817	-0.413	-0.700	-1.352	0.327
Chicago	-0.942	-0.350	0.977	-1.139	0.423	1.112	0.431	-0.201	-1.142
Denver	-0.899	-0.397	-0.820	-0.498	0.661	-0.100	-0.640	-0.611	1.451
Houston	-1.500	-0.789	-0.856	-0.239	-1.419	0.347	-0.470	-0.995	1.848
LA	1.356	0.774	0.636	0.652	-1.585	0.077	0.347	0.640	-0.995
Miami	-0.198	-0.596	-0.941	1.617	-1.078	-1.046	-0.879	0.911	-0.301
NYC	-0.174	0.580	2.135	1.692	0.945	-0.673	2.520	0.355	-0.966
SF	1.511	2.026	-0.156	-0.007	0.752	0.703	-0.264	1.142	-0.006
Seattle	0.879	-0.628	-0.718	-0.876	-0.231	-1.647	-0.568	1.297	-0.220
DC	-0.217	0.719	0.196	-0.591	0.715	1.642	0.225	-1.186	0.005

Conceptual distances: example (2)

- Profile dissimilarities matrix $\boldsymbol{\Delta}$

Atlanta	0.000	3.438	2.036	3.394	4.630	3.930	5.555	4.615	3.330	3.168
Chicago	3.438	0.000	3.635	4.364	3.964	4.740	4.357	4.253	4.299	2.296
Denver	2.036	3.635	0.000	2.333	4.849	3.794	5.674	4.285	3.606	2.994
Houston	3.394	4.364	2.333	0.000	5.016	3.959	6.453	5.517	4.588	3.911
Los Angeles	4.630	3.964	4.849	5.016	0.000	3.361	4.181	3.180	3.611	4.039
Miami	3.930	4.740	3.794	3.959	3.361	0.000	5.234	4.471	2.959	4.906
New York	5.555	4.357	5.674	6.453	4.181	5.234	0.000	4.930	5.535	4.796
San Francisco	4.615	4.253	4.285	5.517	3.180	4.471	4.930	0.000	3.894	3.422
Seattle	3.330	4.299	3.606	4.588	3.611	2.959	5.535	3.894	0.000	4.744
Washington DC	3.168	2.296	2.994	3.911	4.039	4.906	4.796	3.422	4.744	0.000

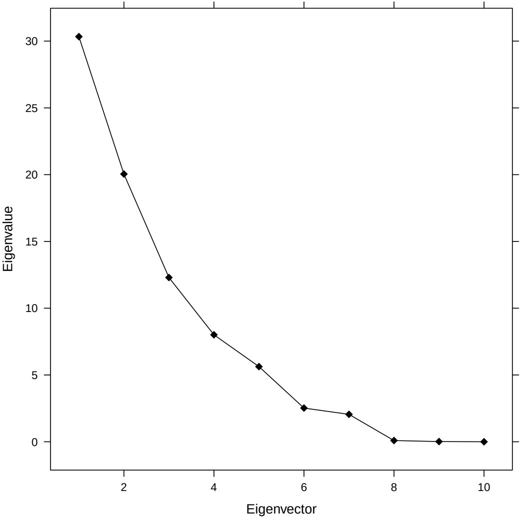
Conceptual distances: example (3)

- Matrix $\mathbf{B}$

Atlanta	5.668	0.028	3.373	2.289	-4.232	-0.963	-4.318	-3.394	0.925	0.623
Chicago	0.028	6.211	-0.889	-1.201	-1.097	-4.205	1.890	-1.515	-2.498	3.277
Denver	3.373	-0.889	5.226	5.104	-5.490	-0.658	-5.212	-2.143	-0.251	0.940
Houston	2.289	-1.201	5.104	10.430	-3.712	1.303	-7.334	-5.579	-1.674	0.374
Los Angeles	-4.232	-1.097	-5.490	-3.712	7.310	1.931	3.191	3.021	0.771	-1.693
Miami	-0.963	-4.205	-0.658	1.303	1.931	7.851	-1.499	-1.644	3.183	-5.298
New York	-4.318	1.890	-5.212	-7.334	3.191	-1.499	16.554	0.548	-3.404	-0.416
San Francisco	-3.394	-1.515	-2.143	-5.579	3.021	-1.644	0.548	8.849	0.477	1.379
Seattle	0.925	-2.498	-0.251	-1.674	0.771	3.183	-3.404	0.477	7.275	-4.805
Washington DC	0.623	3.277	0.940	0.374	-1.693	-5.298	-0.416	1.379	-4.805	5.619

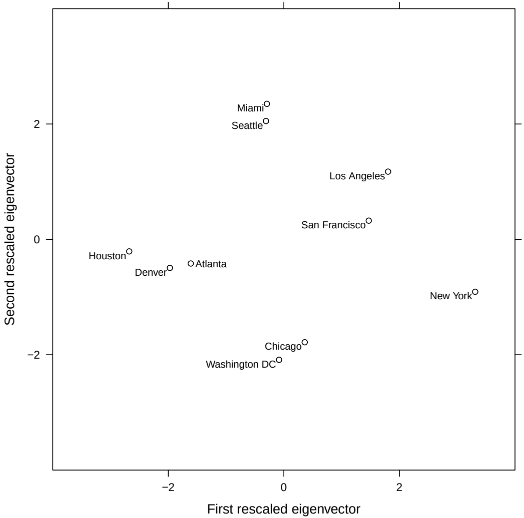
Conceptual distances: example (4)

- No obvious "elbow" in the scree plot.
- No clear distinction between important and unimportant dimensions, in terms of variance explained
- Use $p = 2$ so as to visualize the MDS solution easily.



Conceptual distances: example (5)

- Metric MDS Solution:



Conceptual distances: example (6)

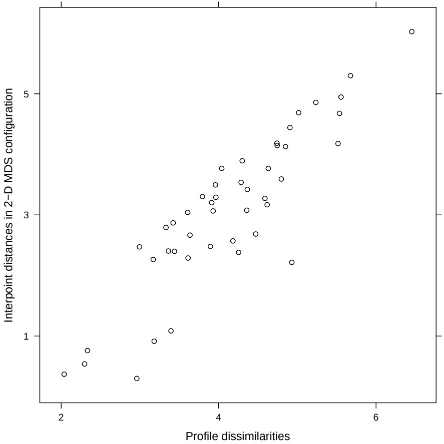
- Calculate fit statistic for the metric MDS solution:

$$\text{Fit} = \frac{\sum_{i=1}^{p} \lambda_i^2}{\sum_{i=1}^{q} \lambda_i^2} = 0.622$$

- The two-dimensional MDS solution explains about 60 percent of the variance in the profile dissimilarities.

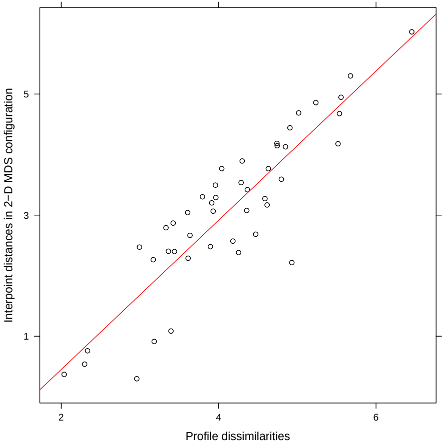
Conceptual distances: example (7) – Shepard Diagram

- Shepard Diagram for metric MDS of city characteristics:



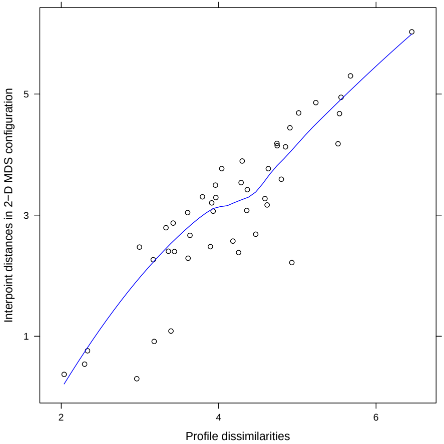
Conceptual distances: example (8) – Shepard Diagram

- Shepard Diagram for metric MDS of city characteristics to test linearity:



Conceptual distances: example (9) – Shepard Diagram

- Shepard Diagram for metric MDS of city characteristics, evidence of non-linearity:



Non-metric MDS

- Distances in MDS solution do not appear to be a linear function of the dissimilarities.
- Instead, distances seem to be monotonically related to dissimilarities.
- Distances are monotonically related to dissimilarities if, for all i, j, and l, the following holds:

$$\delta_{ij} \leq \delta_{il} \Rightarrow d_{ij} \leq d_{il}$$

Distance Scaling

Classical MDS:

- seeks for an optimal configuration $\{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$ that provides

$$\delta_{ij} \approx d_{ij} = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2$$

  as close as possible.

Distance Scaling:

- Relaxing $\delta_{ij} \approx d_{ij}$ by allowing

$$d_{ij} \approx f(\delta_{ij})$$

  where $f(\cdot)$ is a monotonic function
- Called metric MDS if dissimilarities δ_{ij} are quantitative.
- Called non-metric MDS if dissimilarities δ_{ij} are qualitative (e.g. ordinal).
- Unlike classical MDS, distance scaling is an optimization process minimizing stress function, and is solved by iterative algorithms.

Metric MDS

- Given a (low) dimension p and a monotonic function f, the metric MDS seeks for an optimal configuration $\mathbf{X} \subset \mathbb{R}^p$ such that
$$f(\delta_{ij}) \approx d_{ij} = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2$$
as close as possible.
- The function f can be a parametric monotonic function such as $f(\delta_{ij}) = \alpha + \beta\delta_{ij}$
- "As close as possible" means minimizing the loss function
$$stress = L(d_{ij}) = \left(\frac{1}{\sum_{l<k} \delta_{lk}^2} \sum_{i<j} (d_{ij} - f(\delta_{ij}))^2 \right)^{1/2}$$
over d_{ij} and α, β.
- The usual metric MDS is the special case $f(\delta_{ij}) = \delta_{ij}$; its solution (through optimization) = that of the classical MDS.

Sammon mapping

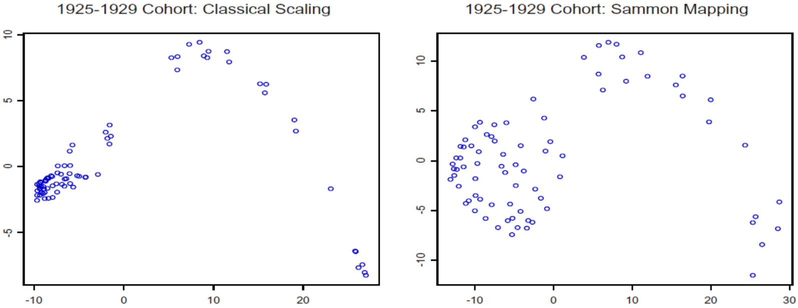
- Sammon mapping is a generalization of the usual metric MDS.
- Sammon's stress (to be minimized) is

$$stress = \frac{1}{\sum_{l<k} \delta_{lk}} \sum_{i<j} \frac{(d_{ij} - \delta_{ij})^2}{\delta_{ij}}$$

- This weighting system normalizes the squared errors in pairwise distances by using the distance in the original space.
- As a result, Sammon mapping preserves the small δ_{ij} better, giving them a greater degree of importance in the fitting procedure than for larger values of δ_{ij}.
- Useful in identifying clusters.
- Optimal solution is found by numerical computation (initial value supplied by the classical MDS).

Classical MDS vs. Sammon Mapping

- Results of cMDS and Sammon mapping for $p = 2$: Sammon mapping better preserves inter-distances for smaller dissimilarities, while proportionally squeezes the inter-distances for larger dissimilarities.



Non-metric MDS

- Often, dissimilarities are known only by their rank order.

Non-metric MDS:

- Given a (low) dimension p and a monotonic function f, the metric MDS seeks for an optimal configuration $\mathbf{X} \subset \mathbb{R}^p$ such that
$$f(\delta_{ij}) \approx d_{ij} = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2$$
as close as possible.
- Unlike the metric MDS, here f is much more general and is only implicitly defined.
- $f(\delta_{ij}) = d_{ij}^*$ are called disparities which only preserve the order of δ_{ij}, i.e.,
$$\delta_{ij} < \delta_{kl} \Leftrightarrow f(\delta_{ij}) \leq f(\delta_{kl}) \Leftrightarrow d_{ij}^* \leq d_{kl}^*$$

Kruskal's non-metric MDS

- Kruskal's non-metric MDS minimizes the stress-1

$$\text{stress-1}(d_{ij}, d^*_{ij}) = \left(\frac{1}{\sum_{l<k} \delta^2_{lk}} \sum_{i<j} \left(d^*_{ij} - \delta_{ij} \right)^2 \right)^{1/2}$$

- the function f works as if it were a regression curve: approximated dissimilarities d_{ij} as observed y, disparities d^*_{ij} as the predicted $\hat{y}$, and the order of dissimilarities as the explanatory variable x.

Kruskal's non-metric MDS (2)

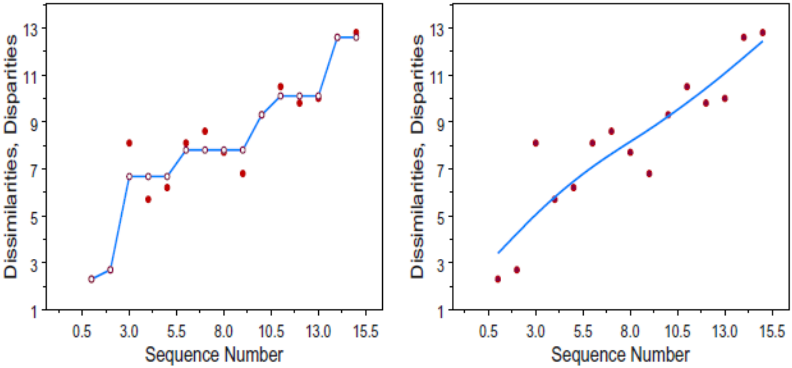


FIGURE 13.10. *Shepard diagram for the artificial example. Left panel: Isotonic regression. Right panel: Monotone spline. Horizontal axis is rank order. For the red points, the vertical axis is the dissimilarity* d_{ij}*, whereas for the fitted blue points, the vertical axis is the disparity* $\widehat{d}_{ij}$.

Example: letter recognition

- Wolford and Hollingsworth (1974) were interested in the confusions made when a person attempts to identify letters of the alphabet viewed for some milliseconds only.
- A confusion matrix shows the frequency with which each stimulus letter was mistakenly called something else.
- A section of this matrix is shown in the table below.

Letter	C	D	G	H	M	N	Q	W
C	–							
D	5	–						
G	12	2	–					
H	2	4	3	–				
M	2	3	2	19	–			
N	2	4	1	18	16	–		
Q	9	20	9	1	2	8	–	
W	1	5	2	5	18	13	4	–

Example: letter recognition (2)

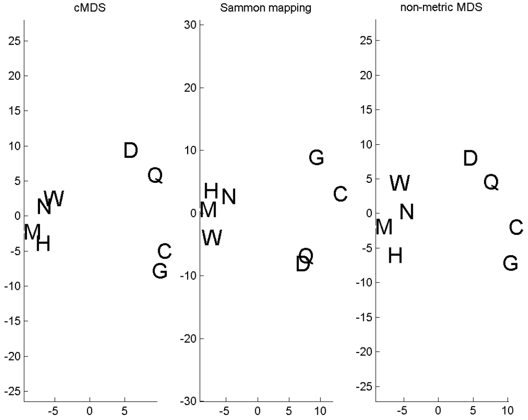
- How to deduce dissimilarities from a similarity matrix? From similarities $\tilde{\delta}_{ij}$, choose a maximum similarity $c \geq \max \tilde{\delta}_{ij}$, so that

$$\delta_{ij} = \begin{cases} c - \tilde{\delta}_{ij}, & i \neq j, \\ 0, & i = j. \end{cases}$$

- The absolute dissimilarities δ_{ij} depend on choice of c. This is the case where the non-metric MDS makes most sense.
- How many dimensions to choose? By inspection of eigenvalues from the classical MDS solution.

Example: letter recognition (3)

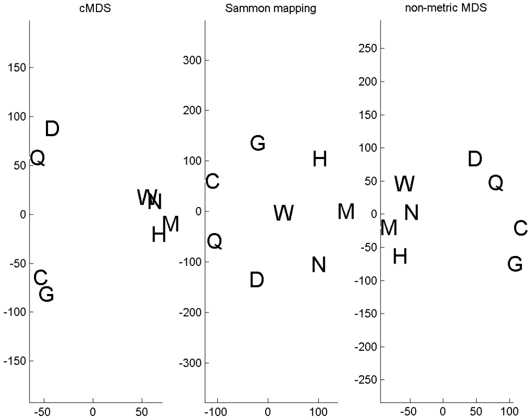
- First choose $c = 21 = \max \delta_{ij} + 1$.
- Compare MDS with $p = 2$: the classical MDS, Sammon mapping, and non-metric scaling (stress-1):



Example: letter recognition (4)

- There are two clusters.
- The eigenvalues of the Gram-matrix $\mathbf{B}$ of the classical MDS are:
```
508.5707
236.0530
124.8229
56.0627
39.7347
-0.0000
-35.5449
-97.1992
```
- The choice of $p = 2$ or $p = 3$ seems reasonable.

- First choose $c = 210 = \max \delta_{ij} + 190$.
- Compare MDS with $p = 2$: the classical MDS, Sammon mapping, and non-metric scaling (stress-1):



Example: letter recognition (6)

- The eigenvalues of the Gram-matrix B of the classical MDS are:

```
1.0e+04 *
2.7210
2.2978
2.1084
1.9623
1.9133
1.7696
1.6842
0.0000
```

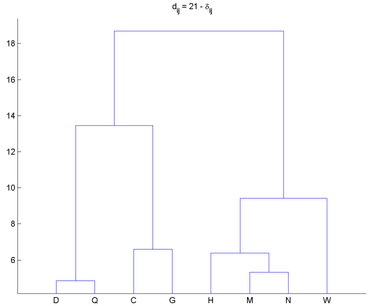
- Need for $p > 3$.
- Sammon mapping fails to reproduce the clusters.

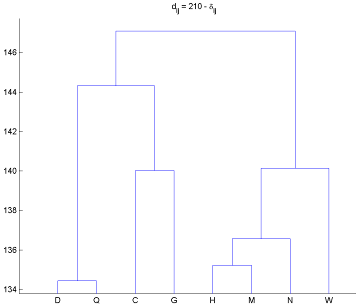
Letter recognition: summary

- The structure of the data appropriate for non-metric MDS.
- Kruskal's non-metric scaling:
  - Appropriate for non-metric dissimilarities (goal is to preserve order)
  - Optimization: susceptible to local minima (leading to different configurations)
  - Time-consuming
- Classical MDS fast, overall good.
- Sammon mapping fails when $c = 210$.

Letter recognition: summary

- Clusters (C; G), (D;Q), (H;M;N;W) are confirmed by a cluster analysis for either choice of c.
- Agglomerative hierarchical clustering with average linkage:





Interpreting an MDS Solution

- Metric and non-metric versions of MDS only determine relative distances between points in scaled m-space.
- The locations of the coordinate axes for the point configuration are completely arbitrary.
  - Final MDS point configuration usually rotated to a "varimax" orientation.
  - Point coordinates usually standardized to a mean of zero on each axis and a variance of 1.0 (or some other specified value).
- Axes have no intrinsic substantive importance or interpretation!

Interpretation Strategies for MDS

- Generally, try to look for two kinds of structure in an MDS solution:
  - Interesting directions within the m-space
    - ★ A direction would usually be "interesting" if points that fall at opposite sides of the space correspond to objects that are contrasting with respect to some substantive characteristic.
  - Interesting groups of points within the m-space:
    - ★ A grouping of points would be "interesting" if the objects corresponding to the grouped points are differentiated from the other objects in terms of some recognizable substantive characteristic.
- Of course, both kinds of structure can occur simultaneously, within a single MDS solution.

Some Cautions About Interpretation

- Simplicity of underlying model and potential for graphical representation of scaling results both facilitate interpretation.
- For many purposes, simply “eyeballing” the point configuration is sufficient for interpretation
- Visual interpretation has potential limitations:
  - It is much more difficult when $m > 2$, and almost impossible when $m > 3$.
  - Highly subjective – we may see structured patterns that are not really there.

For these reasons, it is useful to employ more systematic methods for interpreting an MDS solution

Embedding External Variables

- Researcher often has prior hypotheses about dimensions that differentiate objects in MDS analysis.
- Useful to obtain “external” measures of the objects along these dimensions, separate from the data employed for the MDS.
- If point configuration really does conform to variability of the objects along the external criterion variable, then we can embed an axis representing that dimension within the MDS space.
- Simple regression procedure for doing so.

Embedding External Variables (2)

- Assume an external variable, Y, is available:
  - Each of the n objects in the MDS have scores on the external variable, $y_1, y_2, ..., y_n$.
- Regress Y on the MDS coordinate axes $(\text{Dim}_1, \text{Dim}_2, ..., \text{Dim}_p)$:

$$y_i = \alpha + \beta_1 \text{Dim}_{1i} + \beta_2 \text{Dim}_{2i} + ... + \beta_p \text{Dim}_{pi} + e_i$$

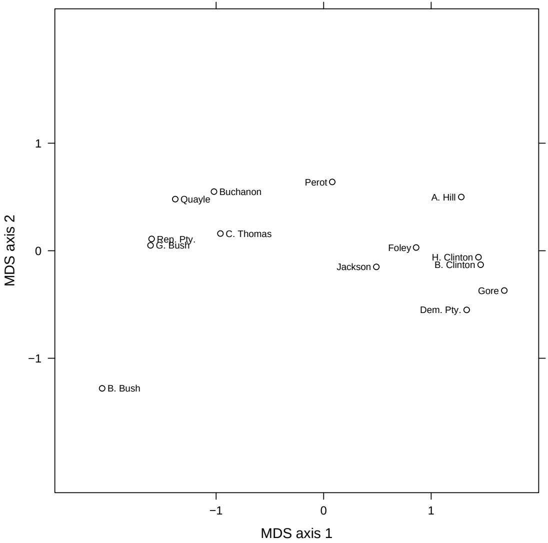
  where Dim_{ij} is the j-th coordinate of the i-th object.
- If regression equation fits well (i.e., R^2 is large), then Y is consistent with the spatial configuration of objects.

Embedding External Variables: data

- A matrix containing perceptual dissimilarities among 14 stimuli, calculated from the 1992 CPS National Election Study

0.00	84.00	47.00	3.00	82.00	73.00	66.00	36.00	81.00	20.00	22.00	54.00	75.00	2.00	George Bush
84.00	0.00	38.00	76.00	9.00	25.00	14.00	91.00	1.00	65.00	68.00	24.00	5.00	86.00	Bill Clinton
47.00	38.00	0.00	40.00	46.00	42.00	26.00	70.00	39.00	32.00	34.00	33.00	44.00	51.00	Ross Perot
3.00	76.00	40.00	0.00	80.00	60.00	56.00	50.00	74.00	18.00	11.00	49.00	79.00	6.00	Dan Quayle
82.00	9.00	46.00	80.00	0.00	29.00	21.00	90.00	12.00	67.00	72.00	41.00	16.00	83.00	Al Gore
73.00	25.00	42.00	60.00	29.00	0.00	13.00	87.00	19.00	64.00	57.00	30.00	31.00	69.00	Anita Hill
66.00	14.00	26.00	56.00	21.00	13.00	0.00	78.00	7.00	43.00	48.00	10.00	23.00	62.00	Thomas Foley
36.00	91.00	70.00	50.00	90.00	87.00	78.00	0.00	89.00	52.00	55.00	77.00	88.00	35.00	Barbara Bush
81.00	1.00	39.00	74.00	12.00	19.00	7.00	89.00	0.00	59.00	63.00	27.00	4.00	85.00	Hillary Clinton
20.00	65.00	32.00	18.00	67.00	64.00	43.00	52.00	59.00	0.00	8.00	37.00	61.00	17.00	Clarence Thomas
22.00	68.00	34.00	11.00	72.00	57.00	48.00	55.00	63.00	8.00	0.00	45.00	58.00	15.00	Pat Buchanon
54.00	24.00	33.00	49.00	41.00	30.00	10.00	77.00	27.00	37.00	45.00	0.00	28.00	53.00	Jesse Jackson
75.00	5.00	44.00	79.00	16.00	31.00	23.00	88.00	4.00	61.00	58.00	28.00	0.00	71.00	Democ. Party
2.00	86.00	51.00	6.00	83.00	69.00	62.00	35.00	85.00	17.00	15.00	53.00	71.00	0.00	Repub. Party

Embedding External Variables: MDS solution



Embedding External Variables: theoretical predictions

- Public perceptions of political figures are affected by two factors:
  - Ideology
  - Overall popularity
- Introducing the two variables:
  - LC: A scale ranging from -100 to 100, with negative values indicating liberal positions, positive values indicating conservative positions.
  - AFF: A 0-100 scale, with larger values corresponding to greater popularity

Embedding External Variables: data matrix with external variables

D1:	D2:	AFFECT:	LC:	NAME:
-1.61	0.05	52	27	G. Bush
1.46	-0.13	56	-22	B. Clinton
0.08	0.64	45	0	Perot
-1.38	0.48	42	29	Quayle
1.68	-0.37	57	-18	Gore
1.28	0.50	49	-19	A. Hill
0.86	0.03	48	-9	Foley
-2.06	-1.28	67	12	B. Bush
1.44	-0.06	54	-17	H. Clinton
-0.96	0.16	45	15	C. Thomas
-1.02	0.55	42	19	Buchanon
0.49	-0.15	47	-16	Jackson
1.33	-0.55	59	-19	Dem. Pty.
-1.60	0.11	52	22	Rep. Pty.

Embedding External Variables: OLS estimates for ideology

- Ideology equation

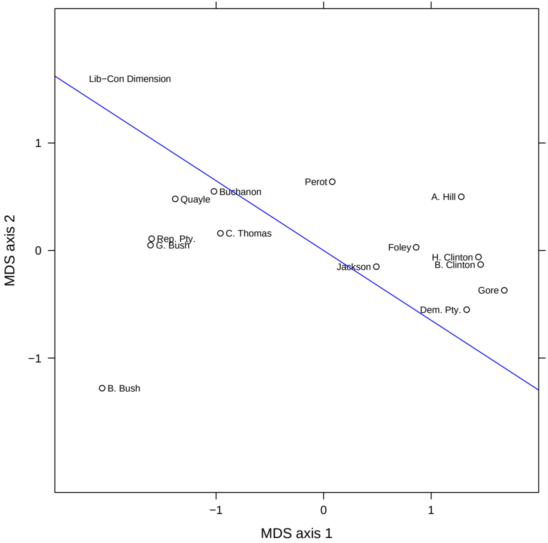
$$\text{LC}_i = 0.289 - 13.343\text{Dim}_{1i} + 8.657\text{Dim}_{2i} + e_i$$

$$R^2 = 0.940$$

$$\text{Slope}_{\text{LC}} = \frac{8.657}{-13.343} = -0.649$$

- Draw line representing ideology dimension into configuration.

Embedding External Variables: inserting ideology



Embedding External Variables: OLS estimates for popularity

- Popularity equation

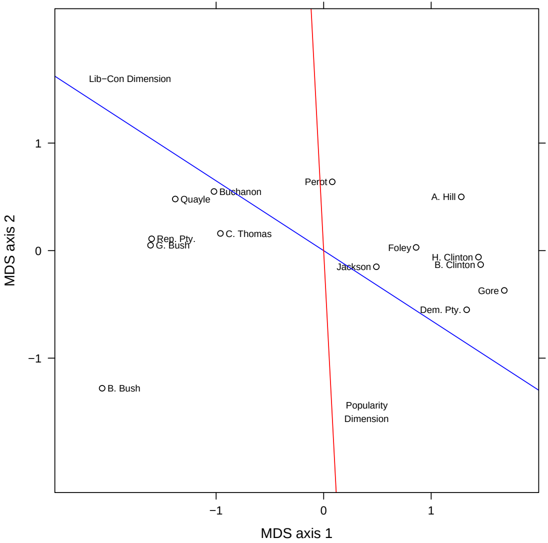
$$\text{AFF}_i = 51.054 + 0.655\text{Dim}_{1i} - 12.622\text{Dim}_{2i} + e_i$$

$$R^2 = 0.832$$

$$\text{Slope}_{\text{AFF}} = \frac{-12.622}{0.655} = -19.270$$

- Draw line representing popularity dimension into configuration.

Embedding External Variables: inserting ideology



Data for MDS

- Direct dissimilarity judgments about stimuli
- Physical distances
- Profile dissimilarities (sum of squared difference measures)
- Confusion measures
- Temporal change rates
- LOS dissimilarities

MDS in R

```
library(MASS)
# compute dissimilarity matrix from a dataset
>d <- dist(swiss)
# d is (n x n-1) lower triangle matrix
>cmdscale(d, k =2) # classical MDS
>sammon(d,k=1) # Sammon Mapping
>isoMDS(d,k=2) # Kruskal's Non-metric MDS
```

References

See Chapter 3 of [1] for more.

F. Husson, S. Le, J. Pagès, Exploratory Multivariate Analysis by Example Using R, Second Edition, Chapman & Hall/CRC Computer Science & Data Analysis, CRC Press, 2017.
URL
https://books.google.com/books?id=nLrODgAAQBAJ