Summary of PCA

e.g. $N = 30$, $P = 200$, $q = 5$

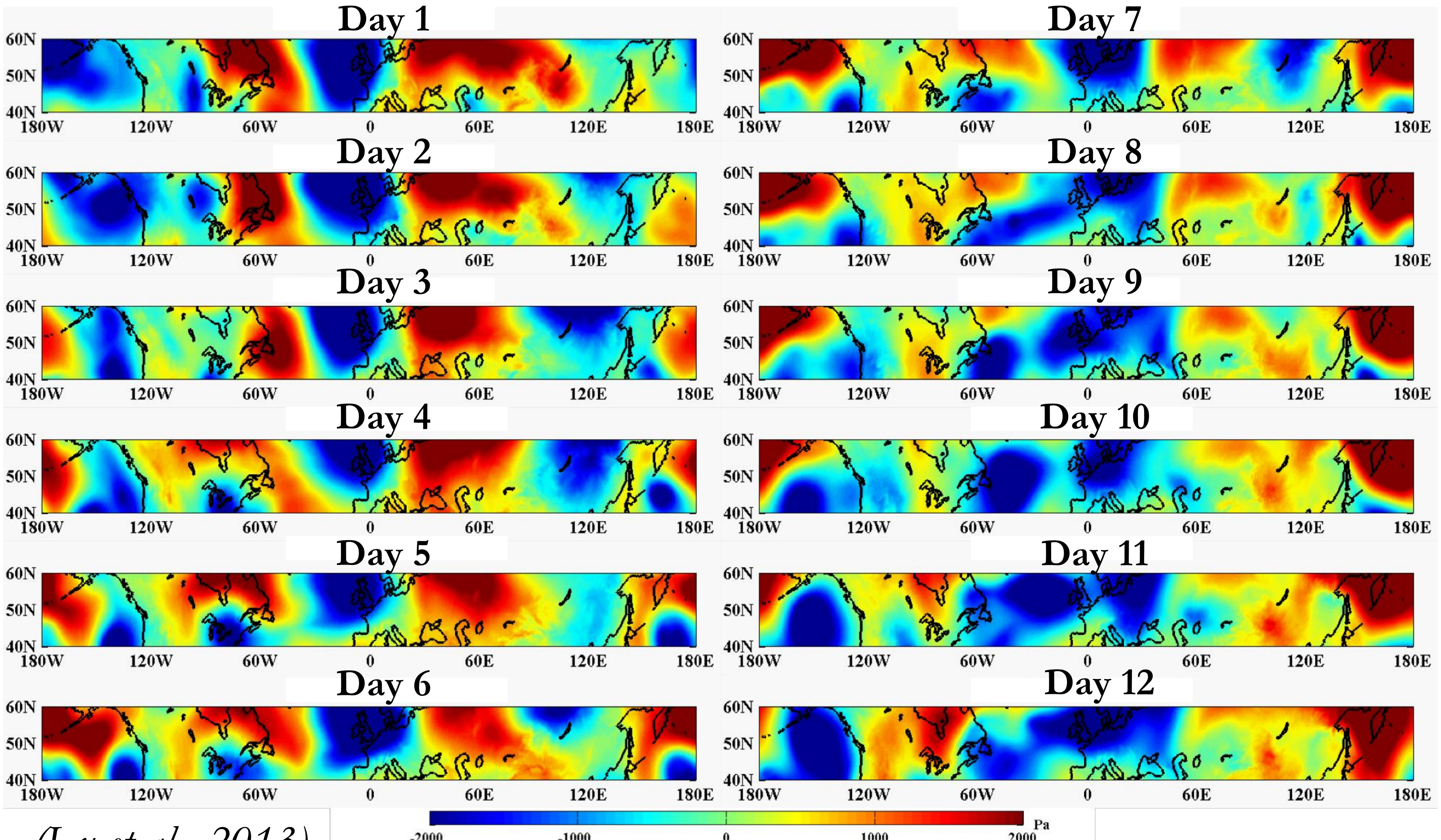
- Input data set: D_{NxP}
- Output (component score matrix): S_{Nxq}

The reduced matrix has scores for each n (=1,…,N) subject on several selected hypothetical variables called PCs.

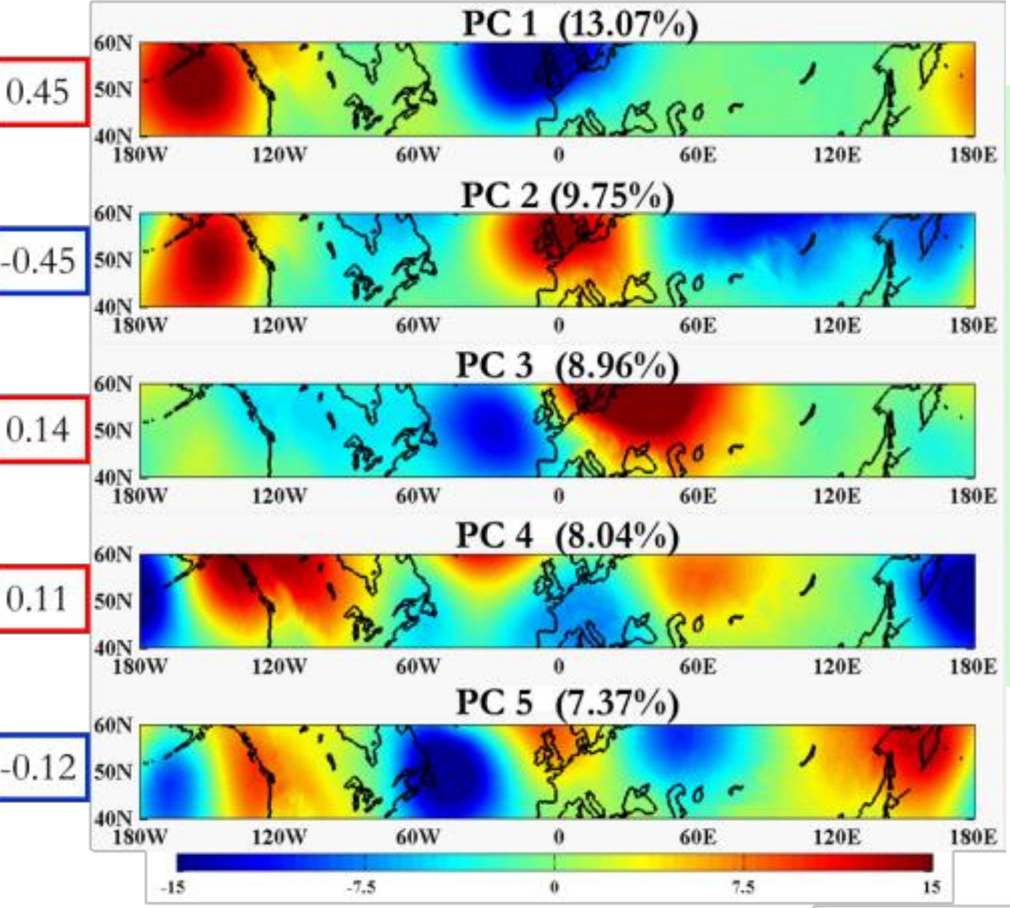
- Why PCA
  1. Different structures overlapped in space or in time or both that needs to be separated

LECTURE WK4B: RECONSTRUCTION

The propagation of the composite wave structure over the extreme rainy days

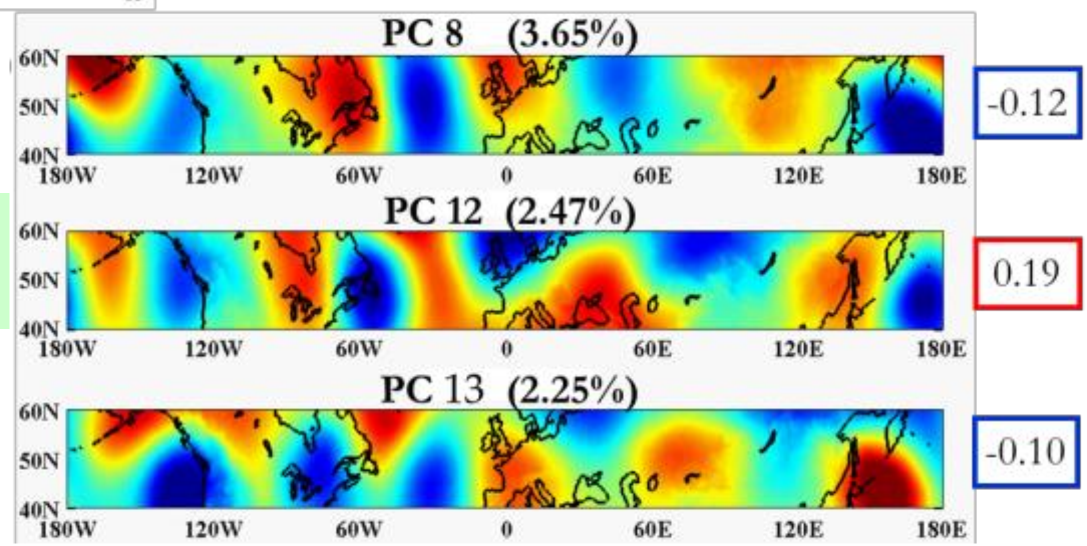


(Lu et al., 2013)



PCA extracts spatial structures and their temporal variations and changes

- Rossby wave structure with different wave numbers
- PC1-5 wavenumber 1-3, potential coupling structures



- PC8, 12, 13 transient structures

(Lu et al., 2013)

Summary of PCA

e.g. N = 30, P = 200, q = 5

- Input data set: D_{NxP}
- Output (component score matrix): S_{Nxq}

The reduced matrix has scores for each n (=1,…,N) subject on several selected hypothetical variables called PCs.

- Why PCA
  1. Raw data has different structures overlapped in space or in time or both that needs to be separated
  2. PCA produces a reduced set of data that maximally captures the information

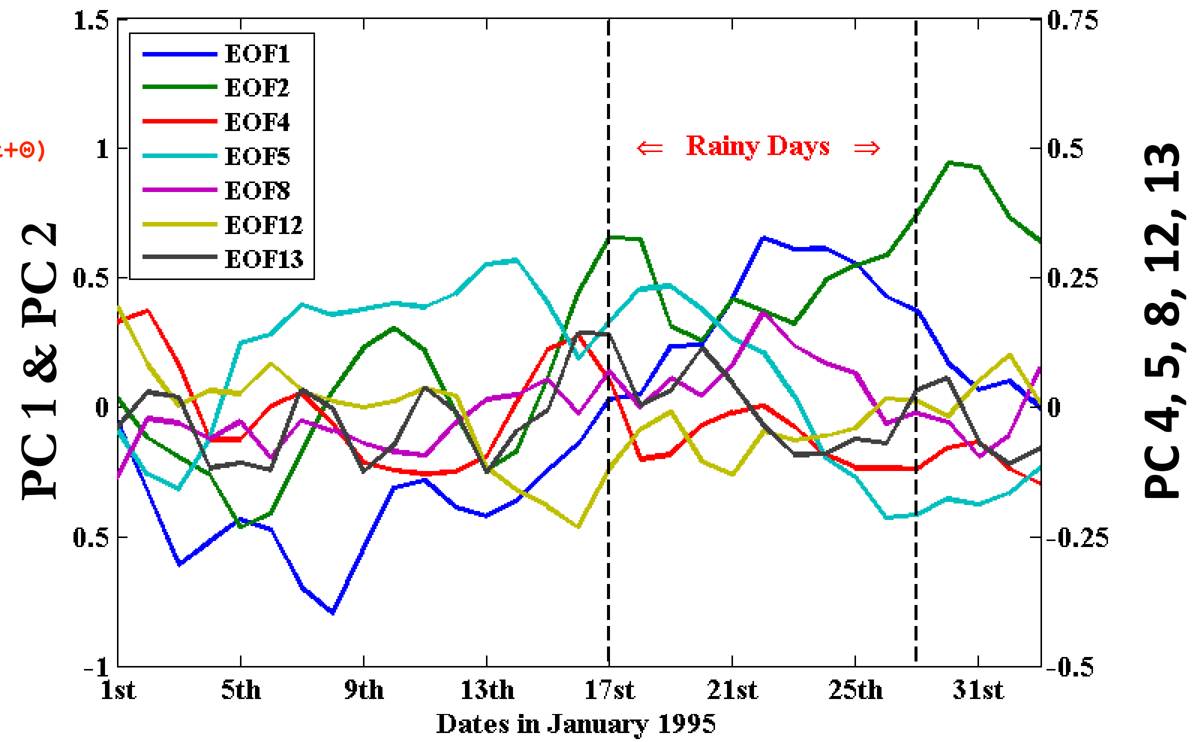
Understand your results

$$D_{N\times P} \approx S_{N\times q} * L_{q\times P}$$

e.g. N = 30, P = 200, q = 5

$L_{q\times P}$: loading matrix – how much each observation "loads" on a particular PC

Spectrum Method:
Decompose a function into ΣA*sin(ωt+Θ)



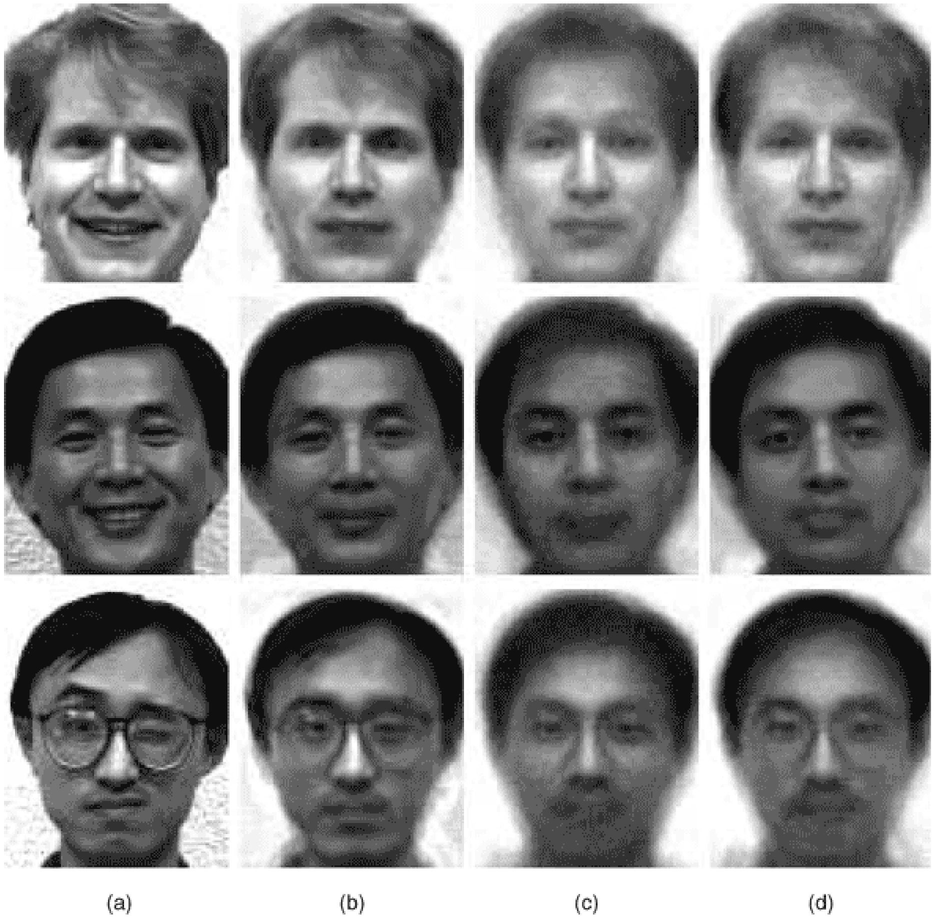
(Lu et al., 2013)

Assess & diagnose by RECONSTRUCTION

$$D_{N \times P} \approx S_{N \times q} * L_{q \times P}$$

- The reconstruction is a multiple regression prediction equation that gives a score based on a weighted summation of other variables; these variables are the hypothetical variables you extracted (components score ***S*** matrix), while the weights are the loadings (loading ***L*** matrix)

Some applications – Image Reconstruction



先线性变换使得眼睛落在同一位置

Nojun Kwak, "Principal Component Analysis Based on L1-Norm Maximization", *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol.30, no. 9, pp. 1672-1680, September 2008, doi:10.1109/TPAMI.2008.114

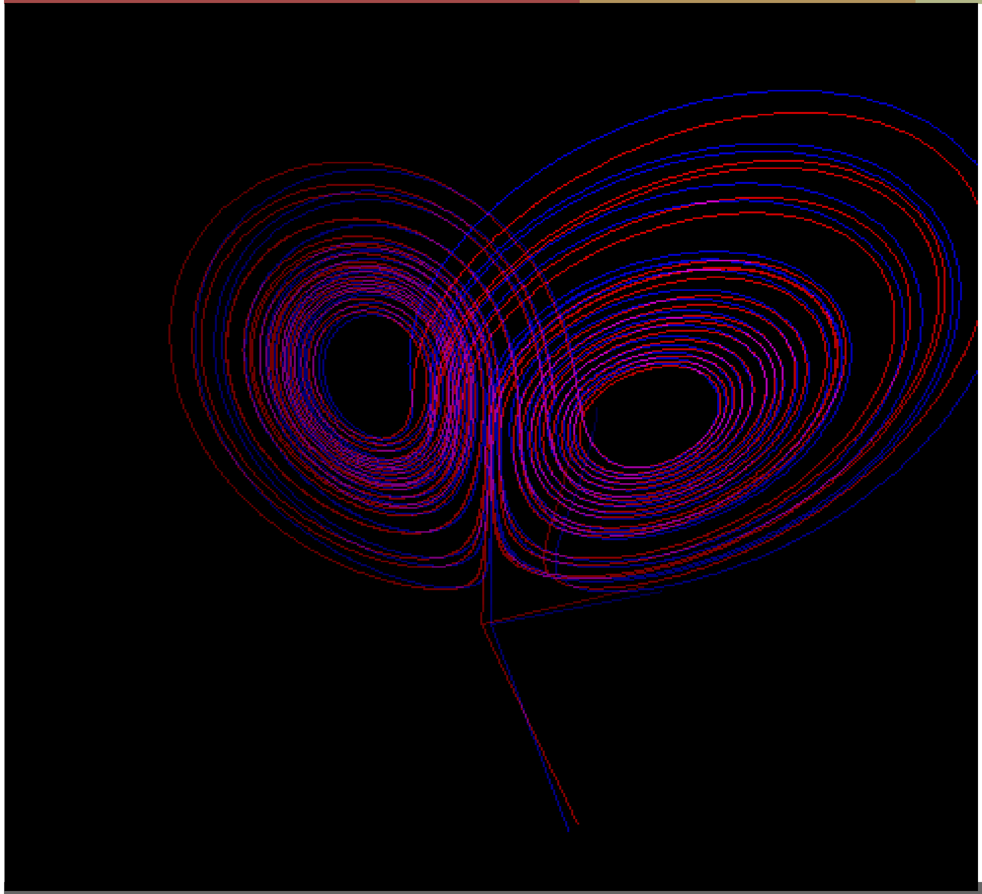
A good example for the Myth: "the more you retain, the better"

You also retain noises when retaining more than necessary principal components

Matthias O. Franz
HTWG Konstanz, Brauneggerstr. 55,
D-78462 Konstanz, Germany

Kwang In Kim & Bernhard Schölkopf
Max-Planck-Institute for Biological Cybernetics,
Spemannstr. 38, D-72076 Tübingen, Germany

MULTIDIMENSIONAL SCALING



Lorenz system by Edward Lorenz

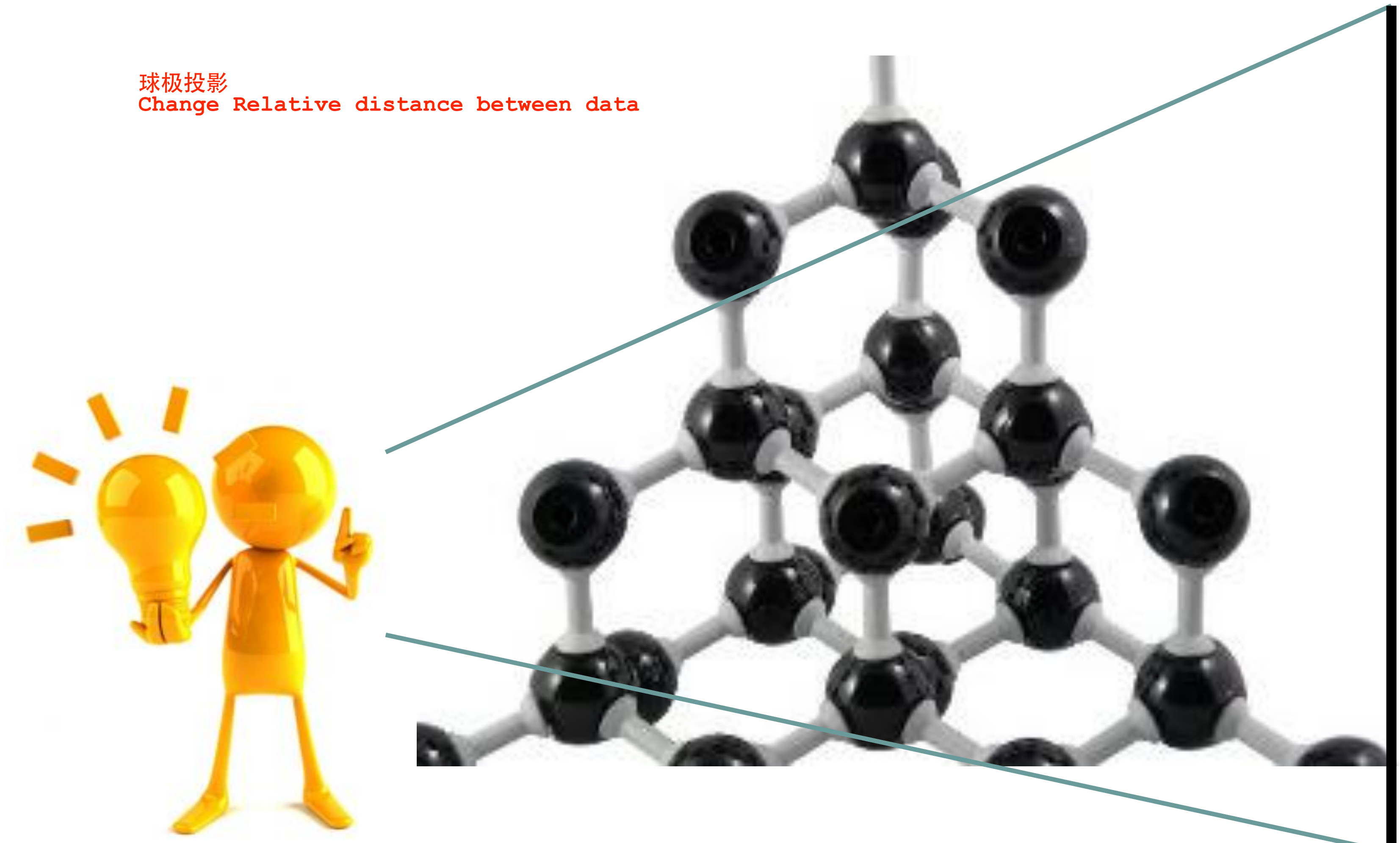
> *"Chaos: When the present determines the future, but the approximate present does not approximately determine the future".*

Danforth, Christopher M. (April 2013). "*Chaos in an Atmosphere Hanging on a Wall*"

Mengqian LU

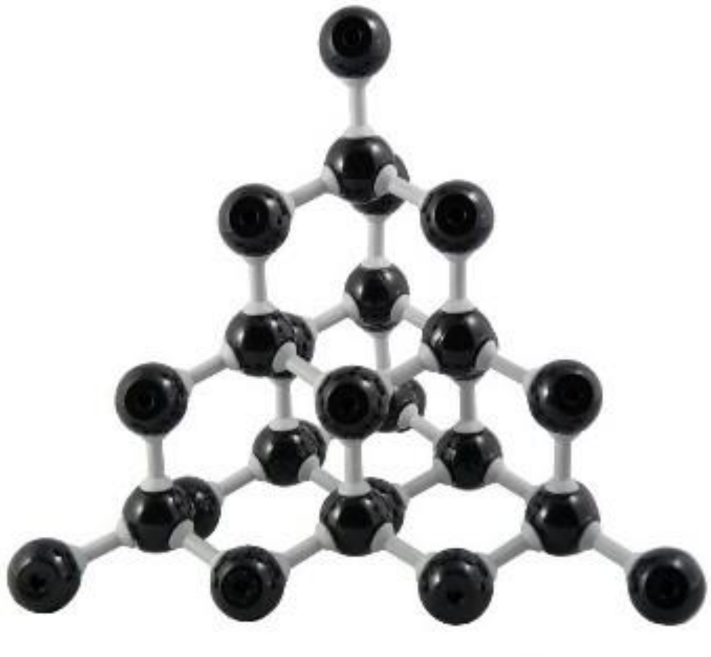
How to represent this in 2-D?

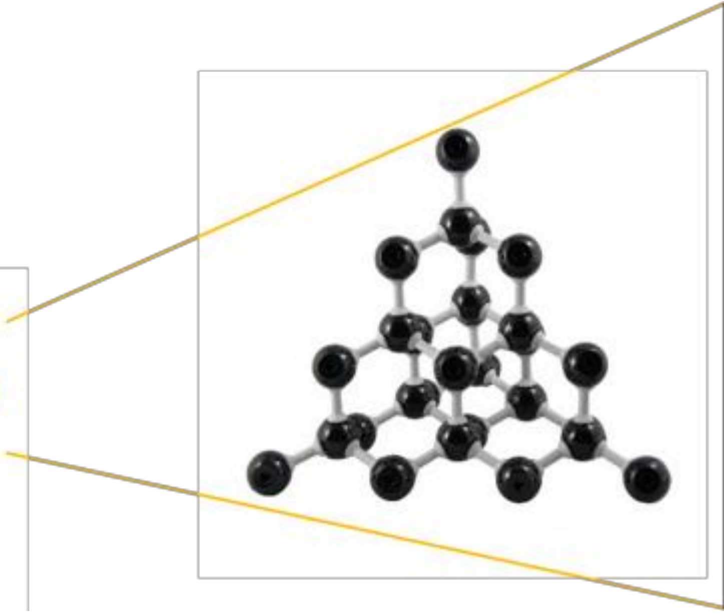
球极投影
Change Relative distance between data



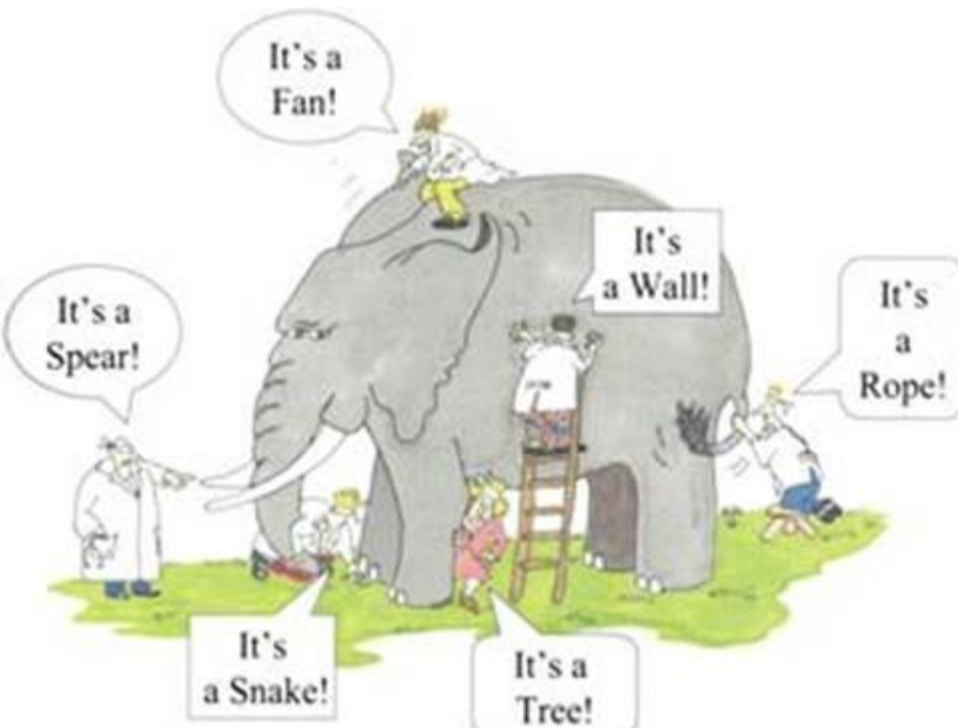
Another idea

Preserve Relative distance between data





MDS

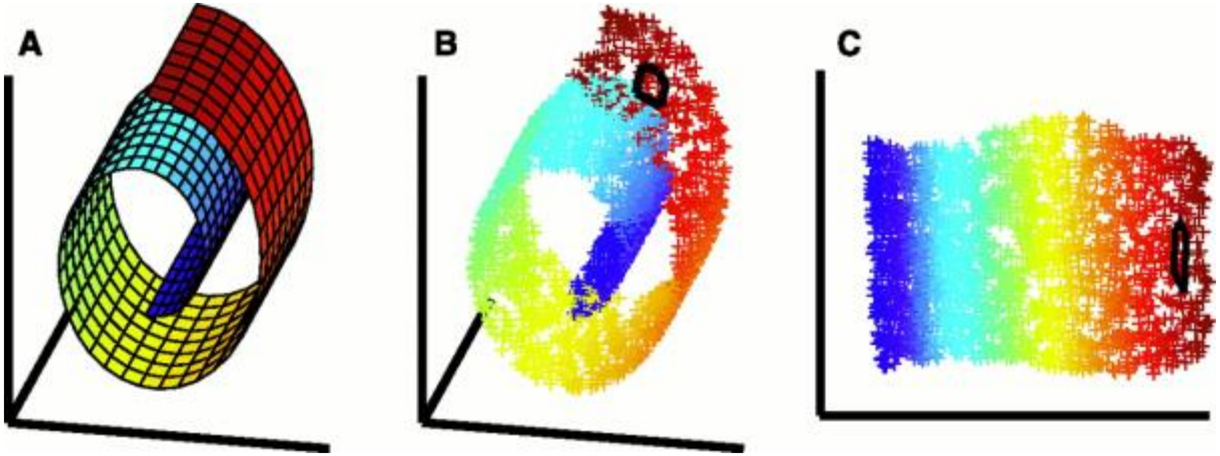


What is MDS? – An Overview

！！！！！MDS 作为聚类分析的头一步　　　　　like a Gaussian Mixed Model

- Find a mapping such that nearby points on the roll (B) are also adjacent in 2D (c) Keeping "**distances**" between points similar

把煎饼果子还原成蛋饼　Cartesian Product of a polar(ρ Θ)*R



高维度的煎饼果子长什么样?

- The "multi" part means the maps not necessarily to be 1 or 2-D

What MDS does?

- **Problem:** high-dimensional data often lies on an intrinsically low-dimensional manifold

Manifold：拓扑空间（常为带有欧氏范数）

- **Goal:** given high-dimensional data, find a low-dimensional coordinates of the data describing where the points lie on the manifold
- **Usage:**
  - Data visualization
  - Pattern recognition
  - Smooth interpolation

How to do MDS?

- The input for MDS is proximities, often proximity matrix.
- MDS looks for low-dimensional spatial configuration to represent the given similarity of dissimilarity

Matrix defining similarity, talk about how we get the matrix later

- Thus first step: *deriving proximities from your raw data*.

1. Direct:
   - A numerical (dis)similarity value to each pair
   - A ranking of the pairs w.r.t. their similarity
2. Indirect:
   - From confusion data
   - From Correlation matrices

Similarity and Dissimilarity Matrix

If you recall PCA,

$$\underset{(n\times p)}{\mathbf{X}} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \xrightarrow{\text{PCA}} Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1q} \\ \cdots & \cdots & \cdots & y_{2q} \\ \cdots & \cdots & \cdots & \cdots \\ y_{n1} & y_{n1} & \cdots & y_{nq} \end{bmatrix}$$

Now we only have proximity matrix represents the similarity/dissimilarity between rows or columns of **X**, then how can you find **X** or **Y**? – MDS

Most R package works with either similarity or dissimilarity, be careful! – Convert to the right matrix!

Some examples: d_{ij} = constant – s_{ij} ; $d_{ij} = 1/s_{ij}$ – constant;

$d_{ij} = s_{ii} + s_{jj} - 2s_{ij}$

Another one in HW3

Classical MDS

- Input: Distance matrix
- Example: given Euclidean distances among selected European cities

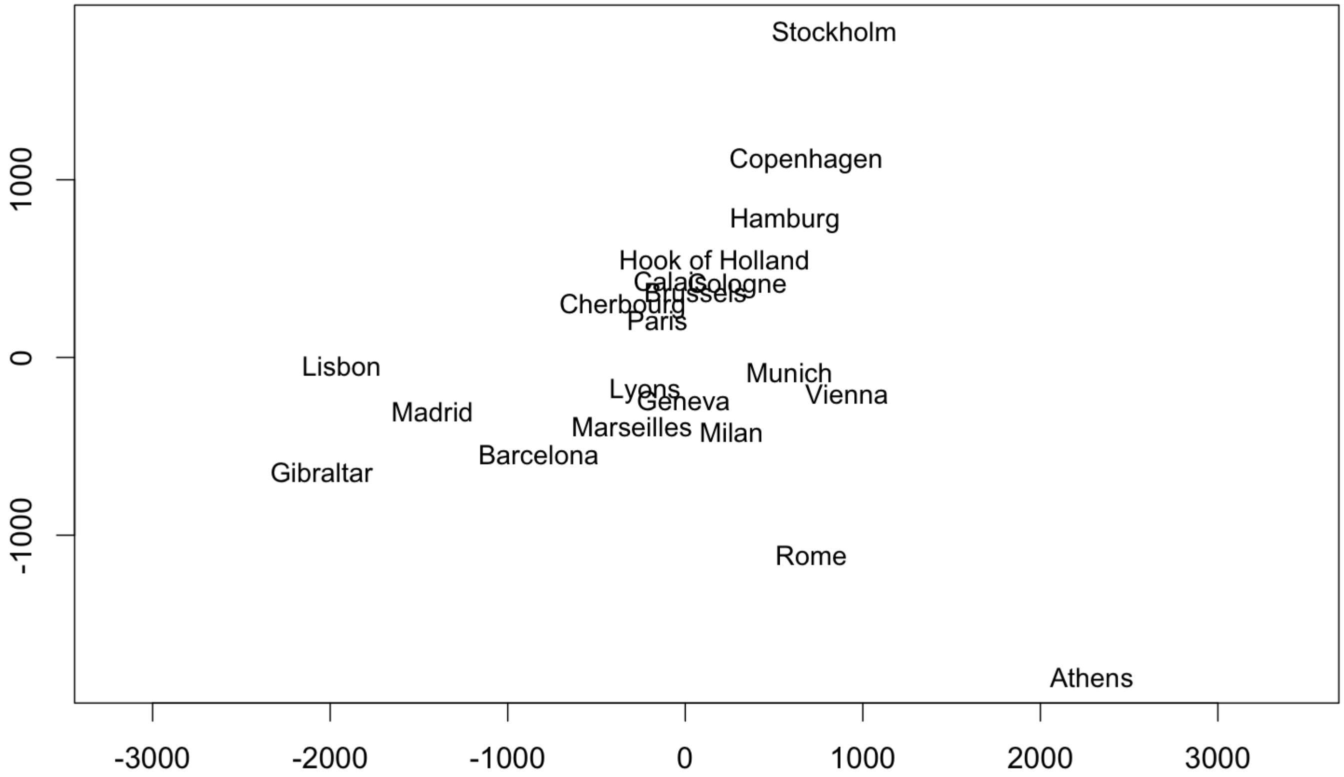
```
> eurodist
```

	Athens	Barcelona	Brussels	Calais	Cherbourg	Cologne	Copenhagen	Geneva	Gibraltar
Barcelona	3313								
Brussels	2963	1318							
Calais	3175	1326	204						
Cherbourg	3339	1294	583	460					
Cologne	2762	1498	206	409	785				
Copenhagen	3276	2218	966	1136	1545	760			
Geneva	2610	803	677	747	853	1662	1418		
Gibraltar	4485	1172	2256	2224	2047	2436	3196	1975	
Hamburg	2977	2018	597	714	1115	460	460	1118	2897
Hook of Holland	3030	1490	172	330	731	269	269	895	2428
Lisbon	4532	1305	2084	2052	1827	2290	2971	1936	676
Lyons	2753	645	690	739	789	714	1458	158	1817
Madrid	3949	636	1558	1550	1347	1764	2498	1439	698
Marseilles	2865	521	1011	1059	1101	1035	1778	425	1693
Milan	2282	1014	925	1077	1209	911	1537	328	2185
Munich	2179	1365	747	977	1160	583	1104	591	2565
Paris	3000	1033	285	280	340	465	1176	513	1971
Rome	817	1460	1511	1662	1794	1497	2050	995	2631
Stockholm	3927	2868	1616	1786	2196	1403	650	2068	3886
Vienna	1991	1802	1175	1381	1588	937	1455	1019	2974

$$\boxed{\{\delta_{ij}\}}$$

R: cmdscale()

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



Invariant under reflection, translation and rotation

Interpretation MDS

- The configuration can be reflected without changing the inter-point distances
- The inter-point distances are not affected if we change the origin
- The set of points (or axes) can be rotated without affecting the inter-point distances

❖ The interpretation of MDS must be invariant under reflection, translation and rotation.

Example: Country Dissimilarities

Kruskal and Wish (1978)

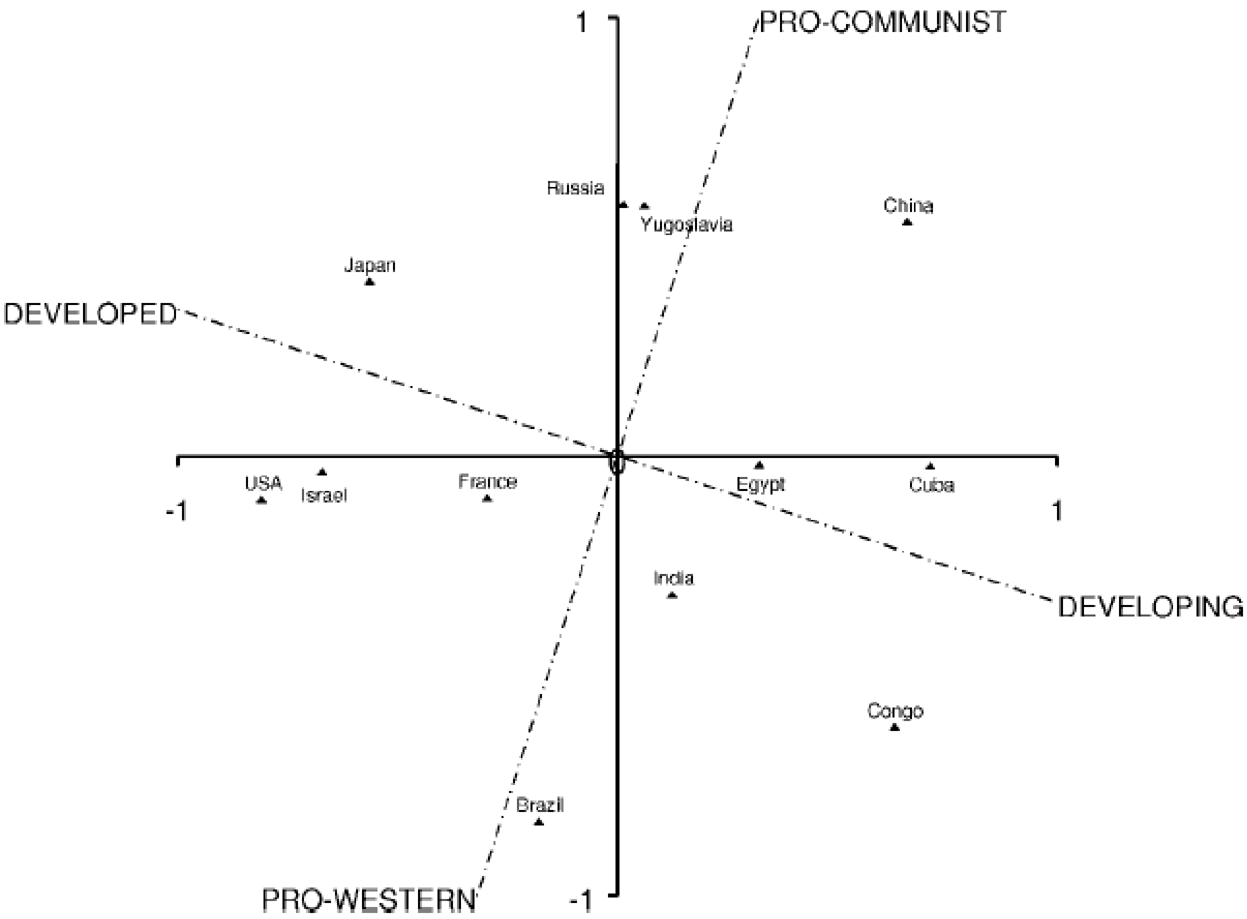
Data from a political science survey: values are average pairwise dissimilarities of countries from a questionnaire given to political science students.

	BEL	BRA	CHI	CUB	EGY	FRA	IND	ISR	USA	USS	YUG
BRA	5.58										
CHI	7.00	6.50									
CUB	7.08	7.00	3.83								
EGY	4.83	5.08	8.17	5.83							
FRA	2.17	5.75	6.67	6.92	4.92						
IND	6.42	5.00	5.58	6.00	4.67	6.42					
ISR	3.42	5.50	6.42	6.42	5.00	3.92	6.17				
USA	2.50	4.92	6.25	7.33	4.50	2.25	6.33	2.75			
USS	6.08	6.67	4.25	2.67	6.00	6.17	6.17	6.92	6.17		
YUG	5.25	6.83	4.50	3.75	5.75	5.42	6.08	5.83	6.67	3.67	
ZAI	4.75	3.00	6.08	6.67	5.00	5.58	4.83	6.17	5.67	6.50	6.92

- Dissimilarity rating: 1 (very different) to 9 (very similar)

Ordinal MDS

Ordinal MDS – Countries' Dissimilarities



Cautionary remarks:

1. Averaged similarities under implicit assumption of no difference among students' subject rating.
   → Individual Scaling or Three-Way Scaling (*Borg and Groenen, 2005* or *Kruskal and Wish, 1978*)

2. Finding the best interpretable axes (rotation) is not the end.
   → Clustering analysis (further)

Some Theory: Classical MDS

- Objective: $\boxed{\{d_{ij}\} \approx (rescaled)\{\delta_{ij}\}}$
- Input: Euclidean distances between n objects in p dimensions
- Output: Coordinates of points invariant to rotation, shift and reflection

- Two steps:
  1. Compute inner product matrix, B, from distance $D = \left\{d_{ij}\right\}$
  2. Compute positions from B

$$d_{ij}^2 = (x_i - x_j)^T (x_i - x_j) = x_i^T x_i + x_j^T x_j - 2x_i^T x_j$$

Let coordinates be $x_i \ (i = 1, \ldots, n)$, where $x_i = (x_{i1}, \ldots, x_{ip})^T$

$$d_{ij}^2 = \sum_{k=1}^{p} (x_{ik} - x_{jk})^2, \text{assuming } \bar{x} = 0$$

$$B = \mathbf{XX}^T, \text{with } b_{ij} = \sum_{k=1}^{p} x_{ik} x_{jk} = x_i^T x_j$$

B: Inner product matrix

$$d_{ij}^2 = b_{ii} + b_{jj} - 2b_{ij}$$

Centering of coordinate matrix X:

$$\sum_{i=1}^{n} b_{ij} = 0$$

$$\sum_{i=1}^{n} b_{ij} = 0 \text{ and } \boxed{d_{ij}^2 = b_{ii} + b_{jj} - 2b_{ij}}$$

$$\sum_i d_{ij}^2 = \sum_i (b_{ii} + b_{jj} - 2b_{ij}) \Leftrightarrow \frac{1}{n}\sum_{i=1}^{n} d_{ij}^2 = \frac{1}{n}\sum_{i=1}^{n} b_{ii} + b_{jj}$$

$$\sum_i d_{ij}^2 = \sum_i (b_{ii} + b_{jj} - 2b_{ij}) \Leftrightarrow \frac{1}{n}\sum_{j=1}^{n} d_{ij}^2 = \frac{1}{n}\sum_{j=1}^{n} b_{jj} + b_{ii}$$

$$\sum_{i,j} d_{ij}^2 = \sum_{i,j} (b_{ii} + b_{jj} - 2b_{ij}) \Leftrightarrow \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}^2 = \frac{2}{n}\sum_{i=1}^{n} b_{ii}$$

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

$B = \mathbf{X}\mathbf{X}^T$ B is a symmetric and positive definite n-by-n matrix

B can be diagonalized: $B = V \Lambda V^T$

$\Lambda = diag(\lambda_1, ..., \lambda_p)$, with $\lambda_1 \geq \lambda_2 ... \geq \lambda_p$ (eigenvalues)

Columns of V are eigenvectors

Some eigenvalues will be zero;

Drop them: $B = V_1 \Lambda_1 V_1^T$

$\mathbf{X} = V_1 \Lambda_1^{\frac{1}{2}}$ "Take square root" of matrix B: $\boxed{\mathbf{X} = V_1 \Lambda_1^{\frac{1}{2}}}$

X contains the point configuration in R^p

Does this remind you of PCA?