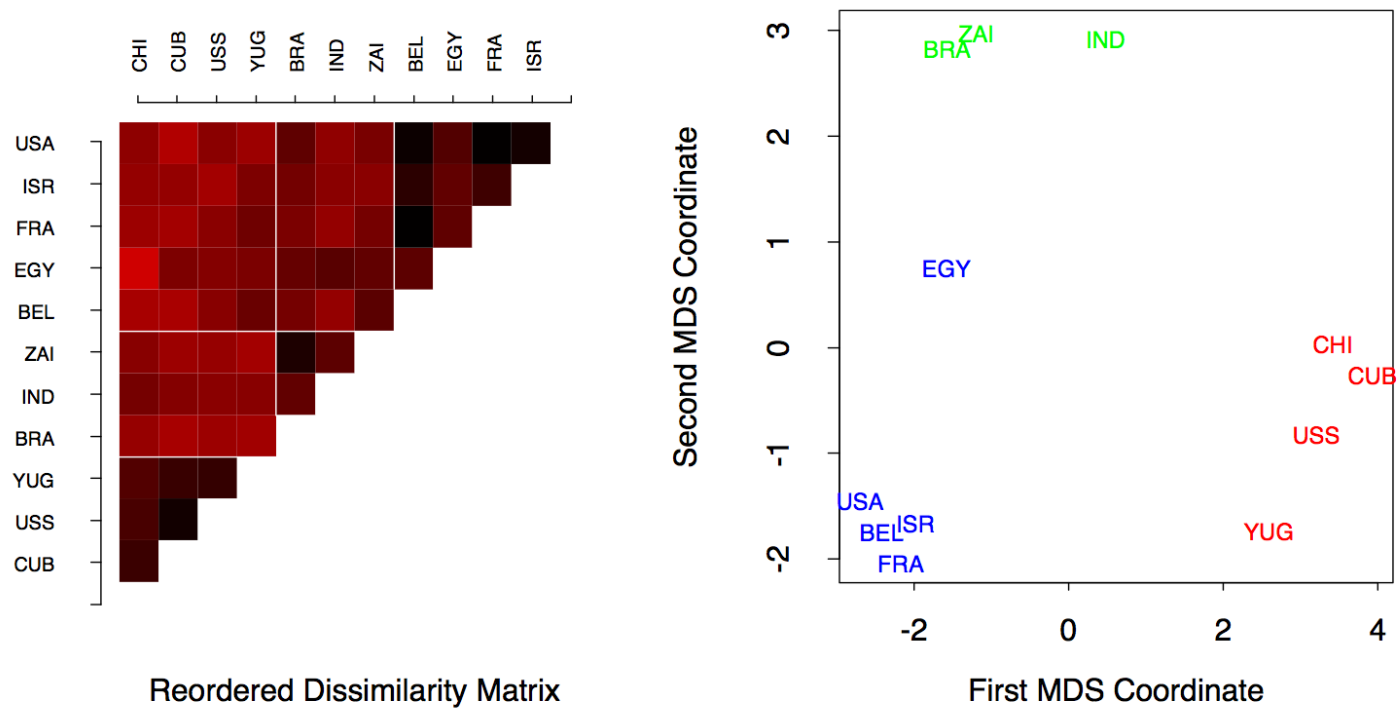
Statistical Machine Learning

Multidimensional scaling

April, 19, 2019

Multidimensional scaling

- Goal of Multidimensional scaling (MDS): Given pairwise dissimilarities, reconstruct a map that preserves distances.
  - From any dissimilarity (no need to be a metric)
  - Reconstructed map has coordinates $x_i = (x_{i1}, x_{i2})$ and the natural distance ($\|x_i - x_j\|_2$)



Multidimensional scaling (cont.)

- MDS is a family of different algorithms, each designed to arrive at optimal low-dimensional configuration ($p = 2$ or 3)
- MDS methods include
  - Classical MDS
  - Metric MDS
  - Non-metric MDS

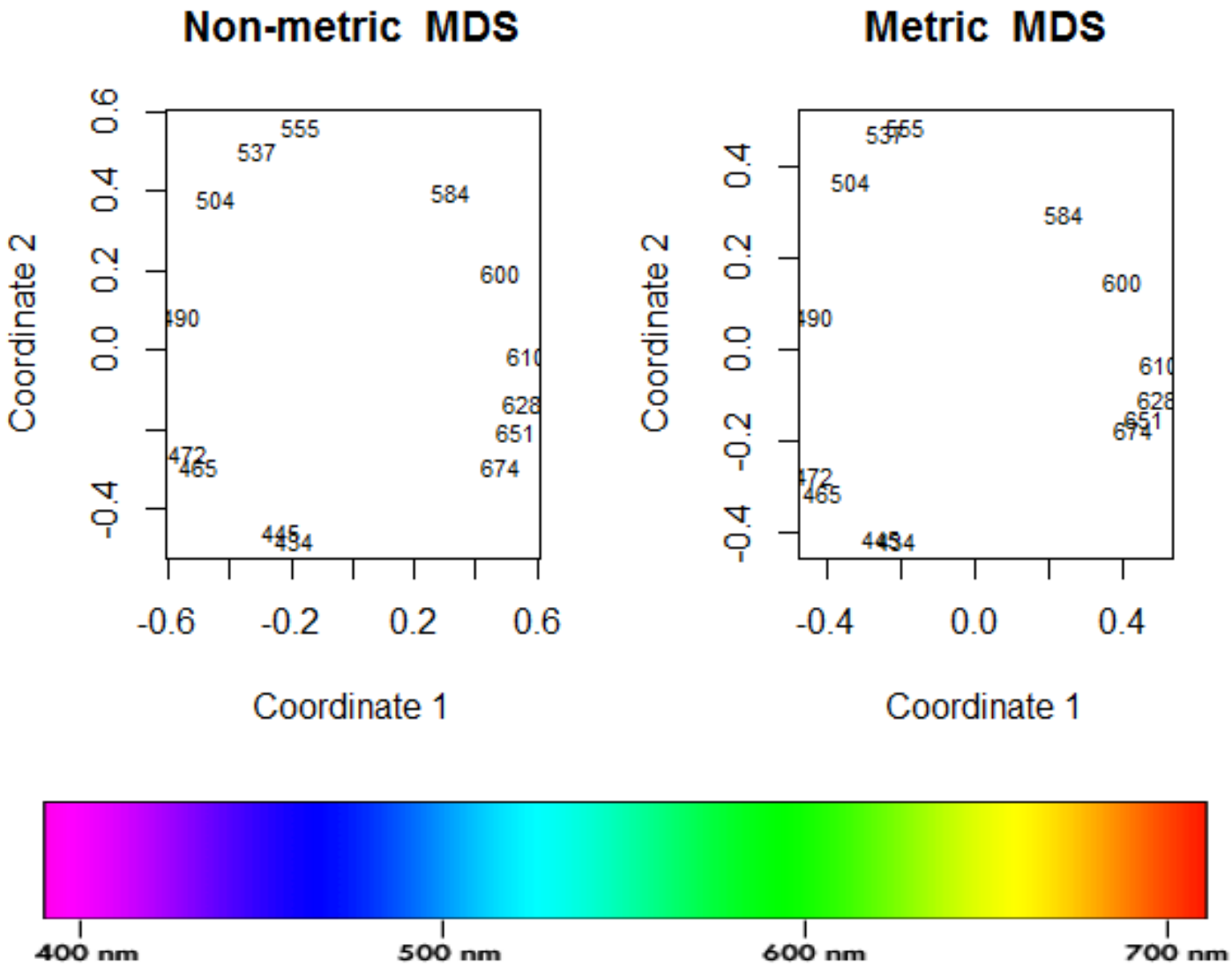
Perception of Color in human vision

- To study the perceptions of color in human vision
- 14 colors differing only in their hue (i.e., wavelengths from 434 μm to 674 μm)
- 31 people to rate on a five-point scale from 0 (no similarity at all) to 4 (identical) for each of the $\binom{14}{2}$ pairs of colors
- Average of 31 ratings for each pair (representing similarity) is then scaled and subtracted from 1 to represent dissimilarities

	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

Perception of Color in human vision (cont.)

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



Distance, dissimilarity and similarity

- Distance, dissimilarity and similarity (or proximity) are defined for any pair of objects in any space.
- In mathematics, a distance function (that gives a distance between two objects) is also called metric, satisfying
  - $d(x, y) \geq 0$
  - $d(x, y) = 0$ if and only if $x = y$
  - $d(x, y) = d(y, x)$
  - $d(x, z) \leq d(x, y) + d(y, z)$
- Given a set of dissimilarities, one can ask whether these values are distances and, moreover, whether they can even be interpreted as Euclidean distances

Euclidean and non-Euclidean distance

- Given a dissimilarity (distance) matrix $D = (d_{ij})$, MDS seeks to find $x_1, \ldots, x_n \in \mathbb{R}^p$ so that

$$d_{ij} \approx \|x_i - x_j\|_2$$

- Oftentimes, for some large p, there exists a configuration $x_1, \ldots, x_n$ with exact distance match $d_{ij} = \|x_i - x_j\|_2$.
- In such a case, the distance d involved is called a Euclidean distance.
- There are, however, cases where the dissimilarity is distance, but there exists no configuration in any p with perfect match
- Such a distance is called non-Euclidean distance. E.g. Radian distance function on a circle.

Classical MDS

- Suppose for now we have euclidean distance matrix $D = (d_{ij})$
- The objective of classical Multidimensional Scaling (cMDS) is to first find $x_1, \ldots, x_n$ so that $\|x_i - x_j\| = d_{ij}$
- It is always possible if the distance is euclidean with x_i's in $n-1$ dimensional space
- Such a solution is not unique, because if $(x_1, \ldots, x_n)$ is the solution, then for any c, $(x_1 + c, \ldots, x_n + c)$ is also a solution.
- Assume that x_i's are centered, i.e., $\sum_{i=1}^{n} x_i = 0$
- If we wish to reduce the dimension, we could perform PCA.

Computation Details

- Let $X = [x_1, \ldots, x_n]$, i.e., each column of X corresponds to one observation.
- Let $B = X^T X \in \mathbb{R}^{n \times n}$ be the gram matrix.
- If the distance is euclidean, it could be shown that

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

- Therefore X could be reconstructed by eigen value decomposition.

Classical MDS Properties

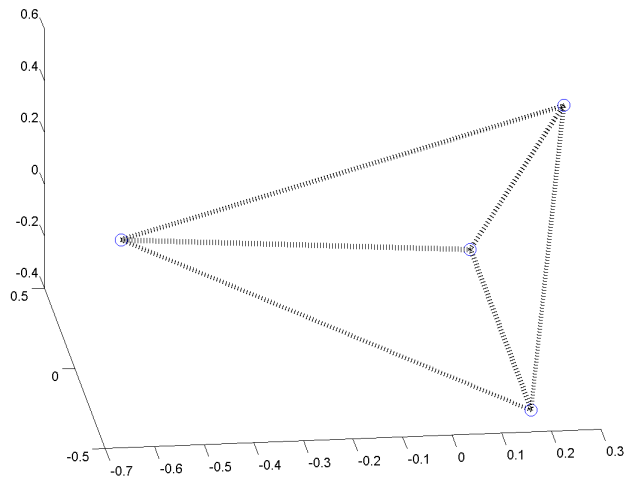
- Assumes that the distance is euclidean
- cMDS gives configurations x_i's for any dimension $p = 1, \ldots, n-1$
- Configuration is centered.
- The coordinates are given by the principal order of largest-to-smallest variances.
- Leads exact solution for Euclidean distances
- Can be used for non-Euclidean distances, in fact, for any dissimilarities. However, it may only able to give configurations for small dimension. In fact, it depends on the number of positive eigen values of B.

cMDS examples: tetrahedron

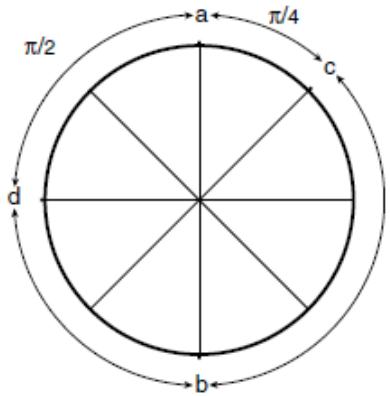
- Pairwise distance matrix for tetrahedron (with distance 1)

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

- Using dimension $p = 3$, we have perfectly retrieved the tetrahedron.

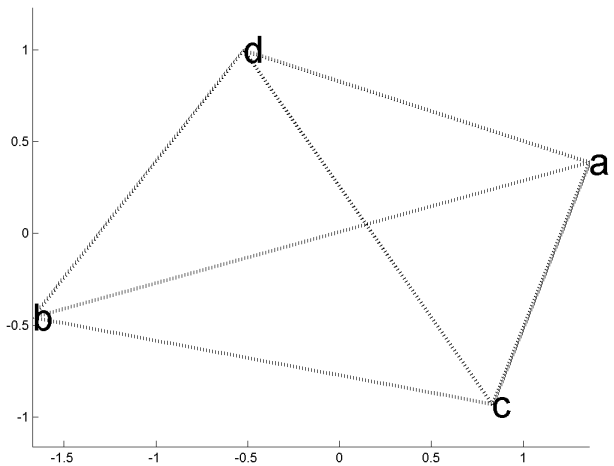


cMDS examples: circular distances



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

- Cannot be embedded in $\mathbb{R}^p$
- The eigenvalues of B are $5.6117, -1.2039, -0.0000, 2.2234$
- Nevertheless, MDS is able to find an optimal configuration. Using dimension $p = 2$



$$\hat{D} = \begin{pmatrix} 0 & 3.1489 & 1.4218 & 1.9784 \\ 3.1489 & 0 & 2.5482 & 1.8557 \\ 1.4218 & 2.5482 & 0 & 2.3563 \\ 1.9784 & 1.8557 & 2.3563 & 0 \end{pmatrix}$$

Distance scaling

- Classical MDS seeks to find an optional configuration x_i that gives $d_{ij} \approx \hat{d}_{ij}$
- Distance Scaling relaxes $d_{ij} \approx \hat{d}_{ij}$ by allowing

$$\hat{d}_{ij} \approx f(d_{ij}), \text{ for some monotone function } f$$

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

The (usual) metric MDS

- Given a (low) dimension p and a monotone function f , metric MDS seeks to find an optimal configuration x_i that gives $d_{ij} \approx \hat{d}_{ij} = \|x_i - x_j\|_2$
- The error is measured by stress

$$\text{stress} = \left(\frac{\sum_{i<j} \{\hat{d}_{ij} - d_{ij}\}^2}{\sum_{i<j} d_{ij}^2} \right)^{1/2}$$

and the metric MDS minimizes the stress over all $\hat{d}_{ij}$.

Sammon mapping

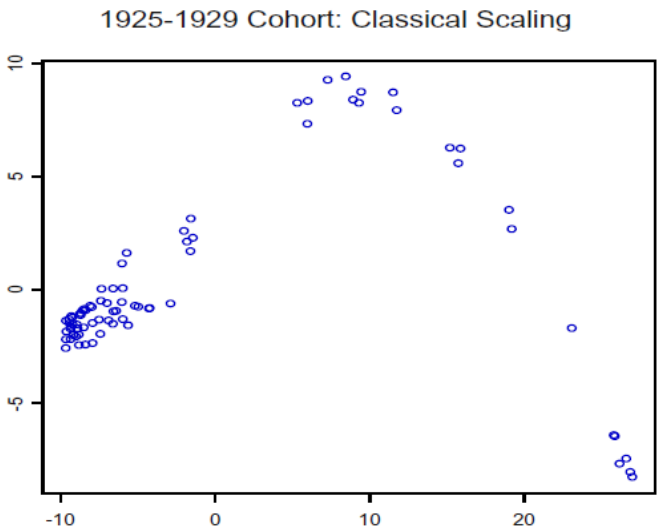
- Sammon mapping is a generalization of the usual metric MDS.
- Sammon's stress (to be minimized) is

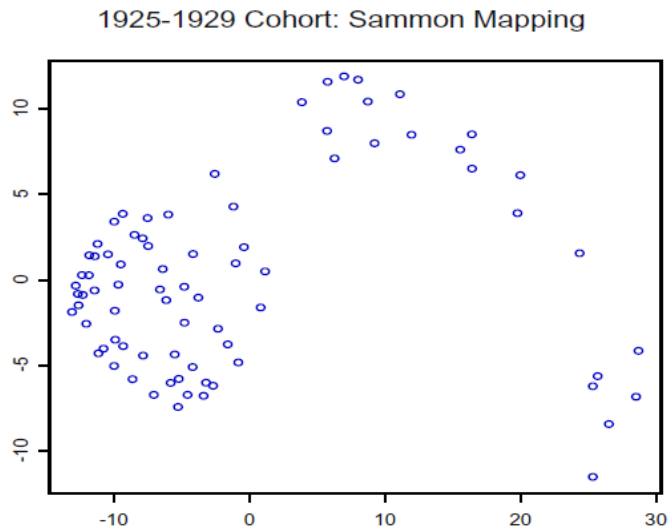
$$\text{Sammon's stress} = \frac{1}{\sum_{i<j} d_{ij}^2} \sum_{i<j} \frac{(\hat{d}_{ij} - d_{ij})^2}{d_{ij}}$$

- 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 d_{ij} , giving them a greater degree of importance in the fitting procedure than for larger values of d_{ij}
- Optimal solution is found by numerical computation (initial value by cMDS).

cMDS 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.





Kruskal's non-metric MDS

- In many applications of MDS, dissimilarities are known only by their rank order, and the spacing between successively ranked dissimilarities is of no interest or is unavailable
- Given a (low) dimension p and a monotone function f, non-metric MDS seeks to find an optimal configuration x_i that gives $d_{ij} \approx f(\hat{d}_{ij}) = f(\|x_i - x_j\|_2)$

$$\text{Kurskal's stress} = \left(\frac{\sum_{i<j} \{f(\hat{d}_{ij}) - d_{ij}\}^2}{\sum_{i<j} d_{ij}^2} \right)^{1/2}$$

- only the order d_{ij} among dissimilarities is needed.