Multidimensional scaling

Wan Nor Arifin

Unit of Biostatistics and Research Methodology, Universiti Sains Malaysia.

email: wnarifin@usm.my

December 16, 2018

Introduction

Clustering

- *Group* subjects/objects that belong to the same group.
- *Clustering* vs *Factoring*?

Clustering

Factoring

- we group variables that represent the same concept/factor.
- analysis on *variables*.

Clustering

- we group subjects/objects/observations that belong to the same group/cluster.
- analysis on *observations*.

Clustering

Orange, motorcycle, bus, durian, banana, car

Group of objects?

Clustering

Group them

[Orange, durian, banana]

[Motorcycle, bus, car]

into two groups

Clustering

Name the group

Fruits	Motor vehicle
Orange	Motorcycle
Durian	Bus
Banana	Car

Cluster/group them in their respective groups.

Multidimensional scaling

MDS

- also known as = perceptual map = spatial map.
- comparison between pairs of objects in term of similarity.
- uncover key dimensions underlying the similarity.
- indentify groups in the suitable dimension.
- raw data – dissimilarity values.

MDS

- analysis on the *dissimilarity* between the objects.
- raw data – starts with dissimilarity matrix.
  - distances between places (e.g. geographical distance between countries)
  - dissimilarity ratings beween objects (e.g. between products).
    - ★ subjective ratings, e.g. on a scale of (similar)1-10(different), rate 10 product pairs.
    - ★ can skip measuring many variables – advantageous in certain research context.
  - based on measured numerical variables, e.g. physical characteristics.
  - turned into Euclidean distances (distance between points in space).

MDS

- analyze the *distances*.
- how similar/dissimilar the object pairs.
  - dissimilar $=\uparrow$ difference $=\uparrow$ distance.
  - similar $=\downarrow$ difference $=\downarrow$ distance.
- find out cluster based on the distances between the objects.

MDS

Basics:

- matrix n x n objects of dissimilarity ratings on object pairs.
- matrix n objects x p variables.

$\rightarrow$ Euclidean distances between n objects.

MDS

Research questions:

- Suitable dimension to visualize the distances?
- How many clusters are there?
- % of variance from the clusters – model fit?

MDS

Applications:

- group products in supermarket or companies in a market survey.
- group diseases based on signs and symptoms.

MDS

Analysis mainly concerned with:

- obtain distances from raw data.
- decide on the number of dimensions – scree-plot of Eigenvalues.
- visualize the clusters in suitable dimension (usually 2D).
- make sense of the clusters and named them.
- % of variance explained by the clusters – $R^2 > 0.6$ for model fit.

Practical

mds.R

Bibliography

Bartholomew, D. J., Steele, F., Moustaki, I., & Galbraith, J. I. (2009). *Analysis of multivariate social science data (2nd eds.)*. Boca Raton, Florida: Chapman & Hall/CRC.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis*. New Jersey: Prentice Hall.