Dimension Reduction Exercise

Ulrike Grömping

06 April 2018

Task 1: Mapping European cities

The R dataset `eurodist` contains road distances for 21 major European cities (not very recent, but very convenient).

i. Create mapping positions for the 21 cities, using classical MDS.
ii. Explore the differences between mapped and original distances (some seem weird!).
iii. Plot the mapped cities (no plot symbols, city names centered on the plot position) without axis distortion (`asp=1`).
iv. Compare the map to the actual geography, and obtain a (hand-crafted) rotation and/or reflection such that mapped locations reflect usual map layout as well as possible (similar to previous task).
v. Create an overlay plot:
    - Draw geo coordinates of the cities (see file provided in moodle) into a coordinate system with lat 30 to 65, lon -15 to 30.
    - Overlay the mapping onto this map.
    - Draw connecting lines between mapped points and the respective geo location for each mapping.

Hints:

- Reflection operations can be done by changing the sign of one or both columns (corresponds to multiplication with an identity matrix with one or both diagonal elements replaced by "-1").
- Clockwise rotation by the angle θ is achieved by multiplication of a 2D row vector **x** with

$$R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

  Multiplying a 2D column vector with that same matrix achieves an anti-clockwise rotation of the same angle.
- Note that your solution may not be the same as someone elses (or mine), because principal component directions (in terms of reflections) may depend on the machine.
- For the scaling in v, rescale scaled versions of the mapping (function `scale`) with the scale information of the geo coordinates that can e.g. be obtained from the attributes "scaled:center" and "scaled:scale" of the result of applying function `scale` .
- The `segments` function can be used for drawing the connecting lines in v.

Task 2: Method exploration with iris data

i. Apply classical MDS to the quantitative variables in the iris data, and interpret the axes. Plot the mapping, and characterize the differences between the three species in terms of the principal axes.
ii. Apply a SOM to the iris data, and inspect its performance graphically.