HDDA Tutorial: MDS

Department of Econometrics and Business Statistics, Monash University

Tutorial 5

For this tutorial we will use the `UScereal` dataset which is available if you install and load the package `MASS`. To load the dataset use the command `data(UScereal)`. Each observation is a brand of breakfast cereal and in total data are available on 11 different variables. Have a look at the help documentation for `UScereal` to familiarise yourself with the data.

1. Remove the non-metric variables `vitamins`, `shelf` and `mfr`.
2. The intention is to use 8-dimensional Euclidean distance between the observations as an input to MDS. Should the data be scaled before computing the distance measure?
3. Find the 2-dimensional classical MDS solution and plot it.
4. Does the plot indicate that one or more cereal brands could be outliers?
5. What are the goodness of fit measures for this solution? Are they same or different?
6. Are there (non-negligible) negative eigenvalues? Why or why not?
7. How would you expect your answer to questions 5 and 6 to change if Manhattan distances are used.
8. Re do the plot but with different coloured labels for each manufacturer. What conclusions do you draw from this analysis?
9. Re-do the analysis using the Sammon mapping. Do your conclusions change?