Bocconi
Università Commerciale Luigi Bocconi

Department of Decision Sciences

Statistics Seminar

How to find the best cluster analysis method (for social stratification based on mixed type data)?

Christian Hennig

Department of Statistical Science, University College London

Thursday, 10 November 2011
12:30pm Room 3-E4-SR03 Via Rontgen 1 Milano

Abstract

This presentation will treat two issues.
1. A general approach for the decision about the “best”' cluster analysis method will be sketched.
This is based on the idea that a subjective decision is needed about the cluster concept required in a particular application, and that there is no such thing as an objectively true clustering determined by the data alone. For example, even if data can be perfectly fitted by a Gaussian mixture, this does not necessarily mean that the mixture components correspond to the ``true'' clusters, because in a given application unimodal mixtures of more than one Gaussian may be considered as a single cluster (see Hennig 2010).
On the other hand, large within-cluster dissimilarities may not be admissible, in which case certain Gaussian components need to be split up.
2. The general philosophy is applied to the problem of defining social strata from data with mixed continuous, ordinal and categorical variables. Clusterings as obtained from a mixture/latent class approach (as for example fitted by the LatentGOLD software, Vermunt and Magidson, 2005) are compared with clusterings obtained from applying k-medoids (Kaufman and Rousseeuw, 1990) to dissimilarities balancing the different types of variables in a sensible user-specified way.

Department of Decision Sciences

Via Röntgen 1 - 20136Milano

Tel. 02 5836.5632
Fax 02 5836.5630