KmL: K-means for Clustering Longitudinal Data

Christophe Genolini[1,*], **Bruno Falissard**[1]

1. INSERM U669, Paris Sud Innovation Group in Adolescent Mental Health Methodology, Paris, France

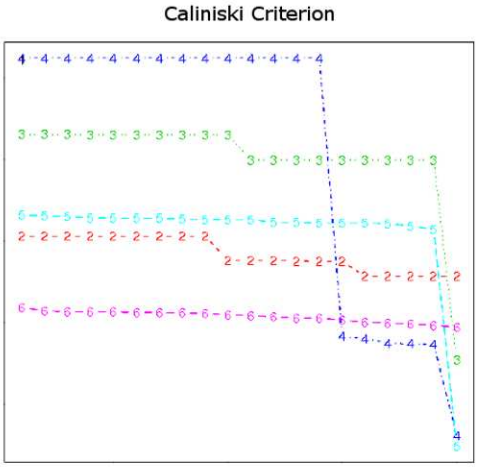
* Contact author : genolini@u-paris10.fr

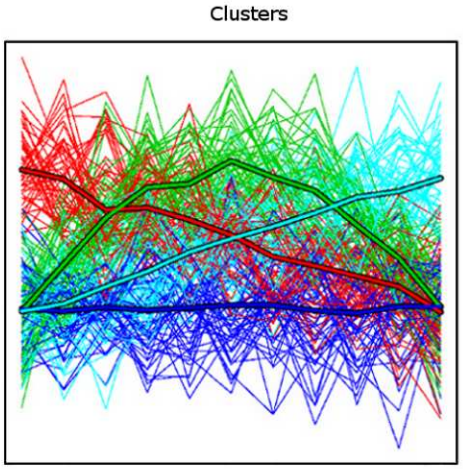
Abstract

The package **KmL**[1] is a generalization of the K-means algorithm for clustering Longitudinal data.

Cohort studies are becoming essential tools in epidemiological research. In these studies, measurements are no longer restricted to a single variable but can be seen as trajectories. K-means is one of the statistical methods that can be used to determine homogeneous groups of patients trajectories.

KmL is a new implementation of k-means design to work specifically with longitudinal data. It provides some facilities to deal with missing values; it also rerolls the algorithm several times, varying the starting conditions and/or the number of clusters looked for; to finish with, its graphical user interface makes this tool well suited to choose the appropriate number of clusters, when the classical critera are not efficient.





Keywords: trajectories, longitudinal data, k-means, cluster analysis, non-parametric algorithm

References

[1] Christophe Genolini, Bruno Falissard. KmL: A Non-Parametric Algorithm for Clustering Longitudinal Data, 2008
http://christophe.genolini.free.fr/kml/.