

Tools and methods for model-based clustering in R

Bettina Grün

Department of Applied Statistics

Johannes Kepler Universität Linz

Altenbergerstraße 69, 4040 Linz, Austria

`Bettina.Gruen@jku.at`

Keywords: clustering, mixture models.

Model-based clustering aims at partitioning observations into groups based on either finite or infinite mixture models. These mixture models differ with respect to their clustering kernel, i.e., the statistical model used for each of the groups. The choice of a suitable clustering kernel allows to adapt the model to the available data structure as well as clustering purpose. We first give an overview on estimation and inference methods for mixture models highlighting the common aspects regardless of the clustering kernel. In addition we consider general tools for assessing and validating the cluster solutions. We then proceed to present the available tools in R implementing these methods. We focus on common structures and implementations which cover a broad range of clustering kernels. In addition general infrastructure packages to allow for joint use of different estimation and inference methods are highlighted and as well as the potential how general tools might be implemented is investigated.