Reign-and-Conquer: Cluster Analysis with a Different Number of Clusters per Margin

Gabriel Martos Venturini

UTDT, Argentina

gmartos@utdt.edu

An often overlooked pitfall of model-based clustering is that it typically results in the same number of clusters per margin, an assumption that may not be natural in practice. We develop a clustering method that takes advantage of the sturdiness of model-based clustering, while attempting to mitigate this issue. The proposed approach allows each margin to have a varying number of clusters and employs a strategy game-inspired algorithm, named "Reign-and-Conquer", to cluster the data. Since the proposed clustering approach only specifies a model for the margins, but leaves the joint unspecified, it has the advantage of being partially parallelizable; hence, the proposed approach is computationally appealing as well as more tractable for moderate to high dimensions than a "full" (joint) model-based clustering approach. A battery of numerical experiments on simulated data indicates an overall good performance of the proposed methods in a variety of scenarios, and real datasets are used to showcase their usefulness in practice.

Trabajo en conjunto con Miguel de Carvalho, The University of Edinburgh, UK y Andrej Svetlosak, The University of Edinburgh, UK.