Better than the best? Answers via model ensemble in density-based clustering. (English)

Summary: With the recent growth in data availability and complexity, and the associated outburst of elaborate modelling approaches, model selection tools have become a lifeline, providing objective criteria to deal with this increasingly challenging landscape. In fact, basing predictions and inference on a single model may be limiting if not harmful; ensemble approaches, which combine different models, have been proposed to overcome the selection step, and proven fruitful especially in the supervised learning framework. Conversely, these approaches have been scantily explored in the unsupervised setting. In this work we focus on the model-based clustering formulation, where a plethora of mixture models, with different number of components and parametrizations, is typically estimated. We propose an ensemble clustering approach that circumvents the single best model paradigm, while improving stability and robustness of the partitions. A new density estimator, being a convex linear combination of the density estimates in the ensemble, is introduced and exploited for group assignment. As opposed to the standard case, where clusters are typically associated to the components of the selected mixture model, we define partitions by borrowing the modal, or nonparametric, formulation of the clustering problem, where groups are linked with high-density regions. Staying in the density-based realm we thus show how blending together parametric and nonparametric approaches may be beneficial from a clustering perspective.

MSC:

62H30 Classification and discrimination; cluster analysis (statistical aspects)
62H99 Multivariate analysis

Keywords:

cluster analysis; model averaging; ensemble learning; density-based clustering; density estimation

Full Text: DOI

References:


Li, J.; Ray, S.; Lindsay, B., A nonparamatric statistical approach to clustering via mode identification, J Mach Learn Res, 8, 1687-1723 (2007) - Zbl 1222.62076


This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.