zbMATH Open
THE FIRST RESOURCE FOR MATHEMATICS

Raftery, Adrian E.; **Dean, Nema**
Variable selection for model-based clustering. (English) Zbl 1118.62339
J. Am. Stat. Assoc. 101, No. 473, 168-178 (2006).

Summary: We consider the problem of variable or feature selection for model-based clustering. The problem of comparing two nested subsets of variables is recast as a model comparison problem and addressed using approximate Bayes factors. A greedy search algorithm is proposed for finding a local optimum in model space. The resulting method selects variables (or features), the number of clusters, and the clustering model simultaneously. We applied the method to several simulated and real examples and found that removing irrelevant variables often improved performance. Compared with methods based on all of the variables, our variable selection method consistently yielded more accurate estimates of the number of groups and lower classification error rates, as well as more parsimonious clustering models and easier visualization of results.

MSC:

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

Cited in **4** Reviews
Cited in **111** Documents

Full Text: DOI Link

Edited by FIZ Karlsruhe, the European Mathematical Society and the Heidelberg Academy of Sciences and Humanities
© 2022 FIZ Karlsruhe GmbH