Model-Based Clustering and Classification for Data Science

With Applications in R

Charles Bouveyron
Université Côte d'Azur

Gilles Celeux
Inria Saclay Île-de-France

T. Brendan Murphy
University College Dublin

Adrian E. Raftery
University of Washington

Cluster analysis finds groups in data automatically. Most methods have been heuristic and leave open such central questions as: how many clusters are there? Which method should I use? How should I handle outliers? Classification assigns new observations to groups given previously classified observations, and also has open questions about parameter tuning, robustness and uncertainty assessment. This book frames cluster analysis and classification in terms of statistical models, thus yielding principled estimation, testing and prediction methods, and sound answers to the central questions. It builds the basic ideas in an accessible but rigorous way, with extensive data examples and R code; describes modern approaches to high-dimensional data and networks; and explains such recent advances as Bayesian regularization, non-Gaussian model-based clustering, cluster merging, variable selection, semi-supervised and robust classification, clustering of functional data, text and images, and co-clustering. Written for advanced undergraduates in data science, as well as researchers and practitioners, it assumes basic knowledge of multivariate calculus, linear algebra, probability and statistics.

August 2019

253 x 177 mm c.446pp 40 b/w illus. 171 colour illus. 48 tables

Hardback 978-1-108-49420-5

Original price	*Discount price*
£59.99	£47.99
$79.99	$63.99

'Bouveyron, Celeux, Murphy, and Raftery pioneered the theory, computation, and application of modern model-based clustering and discriminant analysis. Here they have produced an exhaustive yet accessible text, covering both the field's state of the art as well as its intellectual development. The authors develop a unified vision of cluster analysis, rooted in the theory and computation of mixture models. Embedded R code points the way for applied readers, while graphical displays develop intuition about both model construction and the critical but often-neglected estimation process. Building on a series of running examples, the authors gradually and methodically extend their core insights into a variety of exciting data structures, including networks and functional data. This text will serve as a backbone for graduate study as well as an important reference for applied data scientists interested in working with cutting-edge tools in semi- and unsupervised machine learning.'
John S. Ahlquist,
University of California, San Diego