River Publishers

Elementary Cluster Analysis: Four Basic Methods that (Usually) Work

Author: James C. Bezdek , University of Melbourne (visiting senior fellow), Australia

The availability of packaged clustering programs means that anyone with data can easily do cluster analysis on it. But many users of this technology don't fully appreciate its many hidden dangers. In today's world of "grab and go algorithms," part of my motivation for writing this book is to provide users with a set of cautionary tales about cluster analysis, for it is very much an art as well as a science, and it is easy to stumble if you don't understand its pitfalls. Indeed, it is easy to trip over them even if you do! The parenthetical word usually in the title is very important, because all clustering algorithms can and do fail from time to time.

Modern cluster analysis has become so technically intricate that it is often hard for the beginner or the non-specialist to appreciate and understand its many hidden dangers. Here's how Yogi Berra put it, and he was right:

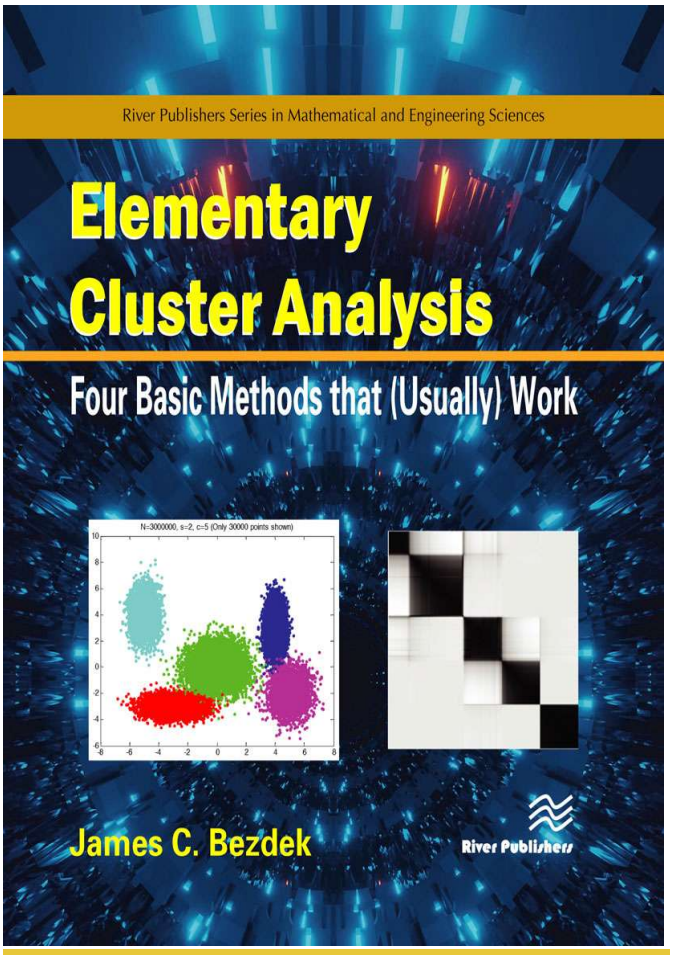
In theory there's no difference between theory and practice. In practice, there is Yogi Berra

This book is a step backwards, to four classical methods for clustering in small, static data sets that have all withstood the tests of time. The youngest of the four methods is now almost 50 years old:

Gaussian Mixture Decomposition (GMD, 1898)
SAHN Clustering (principally single linkage (SL, 1909))
Hard c-means (HCM, 1956, also widely known as (aka) "k-means")
Fuzzy c-means (FCM, 1973, reduces to HCM in a certain limit)

The dates are the first known writing (to me, anyway) about these four models. I am (with apologies to Marvel Comics) very comfortable in calling HCM, FCM, GMD and SL the Fantastic Four.

Cluster analysis is a vast topic. The overall picture in clustering is quite overwhelming, so any attempt to swim at the deep end of the pool in even a very specialized subfield requires a lot of training. But we all start out at the shallow end (or at least that's where we should start!), and this book is aimed squarely at teaching toddlers not to be afraid of the water. There is no section of this book that, if explored in real depth, cannot be expanded into its own volume. So, if your needs are for an in-depth treatment of all the latest developments in any topic in this volume, the best I can do - what I will try to do anyway - is lead you to the pool, and show you where to jump in.



River Publishers Series in Mathematical, Statistical and Computational Modelling for Engineering

ISBN: 9788770224253
e-ISBN: 9788770224246
Available From: March 2022
Price: € 110.00

KEYWORDS:
Cluster Analysis, Unsupervised Learning, k-Means models, Single Linkage, Big Data, Streaming Data

www.riverpublishers.com
marketing@riverpublishers.com