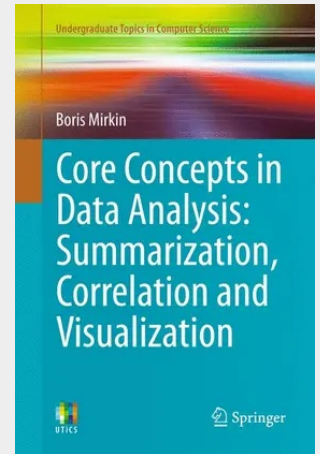


Core Concepts in Data Analysis: Summarization, Correlation and Visualization

Core Concepts in Data Analysis: Summarization, Correlation and Visualization provides in-depth descriptions of those data analysis approaches that either summarize data (principal component analysis and clustering, including hierarchical and network clustering) or correlate different aspects of data (decision trees, linear rules, neuron networks, and Bayes rule). Boris Mirkin takes an unconventional approach and introduces the concept of multivariate data summarization as a counterpart to conventional machine learning prediction schemes, utilizing techniques from statistics, data analysis, data mining, machine learning, computational intelligence, and information retrieval. Innovations following from his in-depth analysis of the models underlying summarization techniques are introduced, and applied to challenging issues such as the number of clusters, mixed scale data standardization, interpretation of the solutions, as well as relations between seemingly unrelated concepts: goodness-of-fit functions for classification trees and data standardization, spectral clustering and additive clustering, correlation and visualization of contingency data. The mathematical detail is encapsulated in the so-called "formulation" parts, whereas most material is delivered through "presentation" parts that explain the methods by applying them to small real-world data sets; concise "computation" parts inform of the algorithmic and coding issues. Four layers of active learning and self-study exercises are provided: worked examples, case studies, projects and questions.

Core Concepts in Data Analysis: Summarization, Correlation and Visualization provides in-depth descriptions of those data analysis approaches that either summarize data (principal component analysis and clustering, including hierarchical and network clustering) or correlate different aspects of data (decision trees, linear rules, neuron networks, and Bayes rule). Boris Mirkin takes an unconventional approach and introduces the concept of multivariate data summarization as a counterpart to conventional machine learning prediction schemes, utilizing techniques from statistics, data analysis, data mining, machine learning, computational intelligence, and information retrieval. Innovations following from his in-depth analysis of the models underlying summarization techniques are introduced, and applied to challenging issues such as the number of clusters, mixed scale data standardization, interpretation of the solutions, as well as relations between seemingly unrelated concepts: goodness-of-fit functions for classification trees and data standardization, spectral clustering and additive clustering, correlation and visualization of contingency data. The mathematical detail is encapsulated in the so-called "formulation" parts, whereas most material is delivered through "presentation" parts that explain the methods by applying them to small real-world data sets; concise "computation" parts inform of the algorithmic and coding issues. Four layers of active learning and self-study exercises are provided: worked examples, case studies, projects and questions.



37,40 €
34,95 € (zzgl. MwSt.)

Nicht mehr lieferbar

Artikelnummer: 9780857292865
Medium: Buch
ISBN: 978-0-85729-286-5
Verlag: Springer
Erscheinungstermin: 07.04.2011
Sprache(n): Englisch
Auflage: 1. Auflage 2011
Serie: Undergraduate Topics in Computer Science
Produktform: Kartoniert
Gewicht: 629 g
Seiten: 390
Format (B x H): 155 x 235 mm

