

1. Record Nr.	UNINA9910337856403321
Autore	Mirkin Boris
Titolo	Core Data Analysis: Summarization, Correlation, and Visualization // by Boris Mirkin
Pubbl/distr/stampa	Cham : , : Springer International Publishing : , : Imprint : Springer, , 2019
ISBN	3-030-00271-3
Edizione	[2nd ed. 2019.]
Descrizione fisica	1 online resource (XV, 524 p. 187 illus., 80 illus. in color.)
Collana	Undergraduate Topics in Computer Science, , 2197-1781
Disciplina	519.50285 519.535
Soggetti	Artificial intelligence - Data processing Data protection Data mining Computer science - Mathematics Data Science Data and Information Security Data Mining and Knowledge Discovery Mathematical Applications in Computer Science
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	Topics in Data Analysis Substance -- Quantitative Summarization -- Learning Correlations -- Core Partitioning: K-Means and Similarity Clustering -- Divisive and Separate Cluster Structures -- Appendix. Basic Math and Code -- Index.
Sommario/riassunto	This text examines the goals of data analysis with respect to enhancing knowledge, and identifies data summarization and correlation analysis as the core issues. Data summarization, both quantitative and categorical, is treated within the encoder-decoder paradigm bringing forward a number of mathematically supported insights into the methods and relations between them. Two Chapters describe methods for categorical summarization: partitioning, divisive clustering and separate cluster finding and another explain the methods for quantitative summarization, Principal Component Analysis and PageRank. Features: · An in-depth presentation of K-means

partitioning including a corresponding Pythagorean decomposition of the data scatter. · Advice regarding such issues as clustering of categorical and mixed scale data, similarity and network data, interpretation aids, anomalous clusters, the number of clusters, etc. · Thorough attention to data-driven modelling including a number of mathematically stated relations between statistical and geometrical concepts including those between goodness-of-fit criteria for decision trees and data standardization, similarity and consensus clustering, modularity clustering and uniform partitioning. New edition highlights: · Inclusion of ranking issues such as Google PageRank, linear stratification and tied rankings median, consensus clustering, semi-average clustering, one-cluster clustering · Restructured to make the logics more straightforward and sections self-contained Core Data Analysis: Summarization, Correlation and Visualization is aimed at those who are eager to participate in developing the field as well as appealing to novices and practitioners. .
