

Data Processing Springer

This book provides a comprehensive picture of mobile big data starting from data sources to mobile data driven applications. Mobile Big Data comprises two main components: an overview of mobile big data, and the case studies based on real-world data recently collected by one of the largest mobile network carriers in China. In the first component, four areas of mobile big data life cycle are surveyed: data source and collection, transmission, computing platform and applications. In the second component, two case studies are provided, based on the signaling data collected in the cellular core network in terms of subscriber privacy evaluation and demand forecasting for network management. These cases respectively give a vivid demonstration of what mobile big data looks like, and how it can be analyzed and mined to generate useful and meaningful information and knowledge. This book targets researchers, practitioners and professors relevant to this field. Advanced-level students studying computer science and electrical engineering will also be interested in this book as supplemental reading.

This textbook, suitable for an early undergraduate up to a graduate course, provides an overview of many basic principles and techniques needed for modern data analysis. In particular, this book was designed and written as preparation for students planning to take rigorous Machine Learning and Data Mining courses. It introduces key conceptual tools necessary for data analysis, including concentration of measure and PAC bounds, cross validation, gradient descent, and principal component analysis. It also surveys basic techniques in supervised (regression and classification) and unsupervised learning (dimensionality reduction and clustering) through an accessible, simplified presentation. Students are recommended to have some background in calculus, probability, and linear algebra. Some familiarity with programming and algorithms is useful to understand advanced topics on computational techniques.

This book has a collection of articles written by Big Data experts to describe some of the cutting-edge methods and applications from their respective areas of interest, and provides the reader with a detailed overview of the field of Big Data Analytics as it is practiced today. The chapters cover technical aspects of key areas that generate and use Big Data such as management and finance; medicine and healthcare; genome, cytochrome and microbiome; graphs and networks; Internet of Things; Big Data standards; bench-marking of systems; and others. In addition to different applications, key algorithmic approaches such as graph partitioning, clustering and finite mixture modelling of high-dimensional data are also covered. The varied collection of themes in this volume introduces the reader to the richness of the emerging field of Big Data Analytics.

This second and revised edition contains a detailed introduction to the key classes of intelligent data analysis methods. The twelve coherently written chapters by leading experts provide complete coverage of the core issues. The first half of the book is devoted to the discussion of classical statistical issues. The following chapters concentrate on machine learning and artificial intelligence, rule induction methods, neural networks, fuzzy logic, and stochastic search methods. The book concludes with a chapter on visualization and an advanced overview of IDA processes.

An Intermediate Course with Examples in S-Plus, R, and SAS
ggplot2

Privacy and Data Protection Issues of Biometric Applications

A Comparative Legal Analysis

Models and Algorithms for Intelligent Data Analysis

Functional and Shape Data Analysis

An Introduction to Data Analysis in R

This book reconsiders statistical methods from the point of view of entropy, and introduces entropy-based approaches for data analysis. Further, it interprets basic statistical methods, such as the chi-square statistic, t-statistic, F-statistic and the maximum likelihood estimation in the context of entropy. In terms of categorical data analysis, the book discusses the entropy correlation coefficient (ECC) and the entropy coefficient of determination (ECD) for measuring association and/or predictive powers in association models, and generalized linear models (GLMs). Through association and GLM frameworks, it also describes ECC and ECD in correlation and regression analyses for continuous random variables. In multivariate statistical analysis, canonical correlation analysis, T2-statistic, and discriminant analysis are discussed in terms of entropy. Moreover, the book explores the efficiency of test procedures in statistical tests of hypotheses using entropy. Lastly, it presents an entropy-based path analysis for structural GLMs, which is applied in factor analysis and latent structure models. Entropy is an important concept for dealing with the uncertainty of systems of random variables and can be applied in statistical methodologies. This book motivates readers, especially young researchers, to address the challenge of new approaches to statistical data analysis and behavior-metric studies.

This book constitutes the refereed conference proceedings of the 9th Annual Privacy Forum, APF 2021. Due to COVID-19 pandemic the conference was held virtually. The 9 revised full papers were carefully reviewed and selected from 43 submissions. The papers are organized in topical sections on Implementing Personal Data Processing Principles; Privacy Enhancing Technologies; Promoting Compliance with the GDPR.

*This textbook offers an accessible introduction to the theory and numerics of approximation methods, combining classical topics of approximation with recent advances in mathematical signal processing, and adopting a constructive approach, in which the development of numerical algorithms for data analysis plays an important role. The following topics are covered: * least-squares approximation and regularization methods * interpolation by algebraic and trigonometric polynomials * basic results on best approximations * Euclidean approximation * Chebyshev approximation * asymptotic concepts: error estimates and convergence rates * signal approximation by Fourier and wavelet methods * kernel-based multivariate approximation * approximation methods in computerized tomography Providing numerous supporting examples, graphical illustrations, and carefully selected exercises, this textbook is suitable for introductory courses, seminars, and distance learning programs on approximation for undergraduate students.*

Students and researchers in the health sciences are faced with greater opportunity and challenge than ever before. The opportunity stems from the explosion in publicly available data that simultaneously informs and inspires new avenues of investigation. The challenge is that the analytic tools required go far beyond the standard methods and models of basic statistics. This textbook aims to equip health care researchers with the most important elements of a modern health analytics toolkit, drawing from the fields of statistics, health econometrics, and data science. This textbook is designed to overcome students' anxiety about data and statistics and to help them to become confident users of appropriate analytic methods for health care research studies. Methods are presented organically, with new material building naturally on what has come before. Each technique is motivated by a topical research question, explained in non-technical terms, and accompanied by engaging explanations and examples. In this way, the authors cultivate a deep ("organic") understanding of a range of analytic techniques, their assumptions and data requirements, and their advantages and limitations. They illustrate all lessons via analyses of real data from a variety of publicly available databases, addressing relevant research questions and comparing findings to those of published studies. Ultimately, this textbook is designed to cultivate health services researchers that are thoughtful and well informed about health data science, rather than data analysts. This textbook differs from the competition in its unique blend of methods and its determination to ensure that readers gain an understanding of how, when, and why to apply them. It provides the public health researcher with a way to think analytically about scientific questions, and it offers well-founded guidance for pairing data with methods for valid analysis. Readers should feel emboldened to tackle analysis of real public datasets using traditional statistical models, health econometrics methods, and even predictive algorithms. Accompanying code and data sets are provided in an author site: <https://roman-gulati.github.io/statistics-for-health-data-science/>

Advances in Signal and Data Processing

Quantitative Data Analysis

Big Data Processing Using Spark in Cloud

With Sample Data, Exercises and Software

Applications

Second EAI International Conference, DIONE 2021, Virtual Event, March 10-12, 2021, Proceedings

Mobile Big Data

With contributions by numerous experts

The fourth edition of this successful textbook presents a comprehensive introduction to statistical and numerical methods for the evaluation of empirical and experimental data. Equal weight is given to statistical theory and practical problems. The concise mathematical treatment of the subject matter is illustrated by many examples and for the present edition a library of Java programs has been developed. It comprises methods of numerical data analysis and graphical representation as well as many example programs and solutions to programming problems. The book is conceived both as an introduction and as a work of reference. In particular it addresses itself to students, scientists and practitioners in science and engineering as a help in the analysis of their data in laboratory courses, in working for bachelor or master degrees, in thesis work, and in research and professional work.

The purpose of this book is to get a practical understanding of the most common processing techniques in earthquake seismology. The book deals with manual methods and computer assisted methods. Each topic will be introduced with the basic theory followed by practical examples and exercises. There are manual exercises entirely based on the printed material of the book, as well as computer exercises based on public domain software. Most exercises are computer based. The software used, as well as all test data are available from <http://extras.springer.com>. This book is intended for everyone processing earthquake data, both in the observatory routine and in connection with research. Using the exercises, the book can also be used as a basis for university courses in earthquake processing. Since the main emphasis is on processing, the theory will only be dealt with to the extent needed to understand the processing steps, however references will be given to where more extensive explanations can be found. Includes: • Exercises • Test data • Public domain software (SEISAN) available from <http://extras.springer.com>

This highly practical handbook is an exhaustive treatment of eddy covariance measurement that will be of keen interest to scientists who are not necessarily specialists in micrometeorology. The chapters cover measuring fluxes using eddy covariance technique, from the tower installation and system dimensioning to data collection, correction and analysis. With a state-of-the-art perspective, the authors examine the latest techniques and address the most up-to-date methods for data processing and quality control. The chapters provide answers to data treatment problems including data filtering, footprint analysis, data gap filling, uncertainty evaluation, and flux separation, among others. The authors cover the application of measurement techniques in different ecosystems such as forest, crops, grassland, wetland, lakes and rivers, and urban areas, highlighting peculiarities, specific practices and methods to be considered. The book also covers what to do when you have all your data, summarizing the objectives of a database as well as using case studies of the CarboEurope and FLUXNET databases to demonstrate the way they should be maintained and managed. Policies for data use, exchange and publication are also discussed and proposed. This one compendium is a valuable source of information on eddy covariance measurement that allows readers to make rational and relevant choices in positioning, dimensioning, installing and maintaining an eddy covariance site; collecting, treating, correcting and analyzing eddy covariance data; and scaling up eddy flux measurements to annual scale and evaluating their uncertainty.

Advances in Data and Information Sciences

Knowledge Graphs and Big Data Processing

Big Data

Hands-on Coding, Data Mining, Visualization and Statistics from Scratch

Statistical Analysis and Data Display

Routine Data Processing in Earthquake Seismology

Analysis of Neural Data

The book describes the emergence of big data technologies and the role of Spark in the entire big data stack. It compares Spark and Hadoop and identifies the shortcomings of Hadoop that have been overcome by Spark. The book mainly focuses on the in-depth architecture of Spark and our understanding of Spark RDDs and how RDD complements big data's immutable nature, and solves it with lazy evaluation, cacheable and type inference. It also addresses advanced topics in Spark, starting with the basics of Scala and the core Spark framework, and exploring Spark data frames, machine learning using MLlib, graph analytics using Graph X and real-time processing with Apache Kafka, AWS Kinesis, and Azure Event Hub. It then goes on to investigate Spark using PySpark and R. Focusing on the current big data stack, the book examines the interaction with current big data tools, with Spark being the core processing layer for all types of data. The book is intended for data engineers and scientists working on massive datasets and big data technologies in the cloud. In addition to industry professionals, it is helpful for aspiring data processing professionals and students working in big data processing and cloud computing environments.

This book is a collection of chapters written by experts on various aspects of big data. The book aims to explain what big data is and how it is stored and used. The book starts from the fundamentals and builds up from there. It is intended to serve as a review of the state-of-the-practice in the field of big data handling. The traditional framework of relational databases can no longer provide appropriate solutions for handling big data and making it available and useful to users scattered around the globe. The study of big data covers a wide range of issues including management of heterogeneous data, big data frameworks, change management, finding patterns in data usage and evolution, data as a service, service-generated data, service management, privacy and security. All of these aspects are touched upon in this book. It also discusses big data applications in different domains. The book will prove useful to students, researchers, and practicing database and networking engineers.

This book presents intelligent data analysis as a tool to fight against COVID-19 pandemic. The intelligent data analysis includes machine learning, natural language processing, and computer vision applications to teach computers to use big data-based models for pattern recognition, explanation, and prediction. These functions are discussed in detail in the book to recognize (diagnose), predict, and explain (treat) COVID-19 infections, and help manage socio-economic impacts. It also discusses primary warnings and alerts; tracking and prediction; data dashboards; diagnosis and prognosis; treatments and cures; and social control by the use of intelligent data analysis. It provides analysis reports, solutions using real-time data, and solution through web applications details.

This open access book is a step-by-step introduction on how shell scripting can help solve many of the data processing tasks that Health and Life specialists face everyday with minimal software dependencies. The examples presented in the book show how simple command line tools can be used and combined to retrieve data and text from web resources, to filter and mine literature, and to explore the semantics encoded in biomedical ontologies. To store data this book relies on open standard text file formats, such as TSV, CSV, XML, and OWL, that can be open by any text editor or spreadsheet application. The first two chapters, Introduction and Resources, provide a brief introduction to the shell scripting and describe popular data resources in Health and Life Sciences. The third chapter, Data Retrieval, starts by introducing a common data processing task that involves multiple data resources. Then, this chapter explains how to automate each step of that task by introducing the required commands line tools one by one. The fourth chapter, Text Processing, shows how to filter and analyze text by using simple string matching techniques and regular expressions. The last chapter, Semantic Processing, shows how XPath queries and shell scripting is able to process complex data, such as the graphs used to specify ontologies. Besides being almost immutable for more than four decades and being available in most of our personal computers, shell scripting is relatively easy to learn by Health and Life specialists as a sequence of independent commands. Comprehending them is like conducting a new laboratory protocol by testing and understanding its procedural steps and variables, and combining their intermediate results. Thus, this book is particularly relevant to Health and Life specialists or students that want to easily learn how to process data and text, and which in return may facilitate and inspire them to acquire deeper bioinformatics skills in the future.

How to Create Value with Artificial Intelligence

An Introduction

Artificial Intelligence in Data and Big Data Processing

Applications to Music Information Retrieval and Psychophysiology of Hearing

9th Annual Privacy Forum, APF 2021, Oslo, Norway, June 17 – 18, 2021, Proceedings

A Companion for Accounting and Information Systems Research

Topological Data Analysis

This book presents the select peer-reviewed proceedings of the International Conference on Signal and Data Processing (ICSDP) 2019. It examines and deliberates on the recent progresses in the areas of communication and signal processing. The book includes topics on the recent advances in the areas of wired and wireless communication, low complexity architecture of MIMO receivers, applications on wireless sensor networks and internet of things, signal processing, image processing and computer vision, VLSI embedded systems, cognitive networks, power electronics and automation, mechatronics based applications, systems and control, cognitive science and machine intelligence, information security and big data. The contents of this book will be useful for beginners, researchers, and professionals interested in the area of communication, signal processing, and allied fields.

This book covers cutting-edge and advanced research on data processing techniques and applications for Cyber-Physical Systems. Gathering the proceedings of the International Conference on Data Processing Techniques and Applications for Cyber-Physical Systems (DPTA 2019), held in Shanghai, China on November 15–16, 2019, it examines a wide range of topics, including: distributed processing for sensor data in CPS networks; approximate reasoning and pattern recognition for CPS networks; data platforms for efficient integration with CPS networks; and data security and privacy in CPS networks. Outlining promising future research directions, the book offers a valuable resource for students, researchers and professionals alike, while also providing a useful reference guide for newcomers to the field.

This book constitutes the refereed post-conference proceedings of the Second International Conference on Data Information in Online Environments, DIONE 2021, which took place in March 2021. Due to COVID-19 pandemic the conference was held virtually. DIONE 2021 presents theoretical proposals and practical solutions in the treatment, processing and study of data and information produced in online environments, the latest trends in the analysis of network information, media metrics social, data processing technologies and open science. The 40 revised full papers were carefully reviewed and selected from 86 submissions. The papers are grouped in thematical sessions on evaluation of science in social networking environment; scholarly publishing and online communication; and education in online environments.

This book discusses all critical privacy and data protection aspects of biometric systems from a legal perspective. It contains a systematic and complete analysis of the many issues raised by these systems based on examples worldwide and provides several recommendations for a transnational regulatory framework. An appropriate legal framework is in most countries not yet in place. Biometric systems use facial images, fingerprints, iris and/or voice in an automated way to identify or to verify (identity) claims of persons. The treatise which has an interdisciplinary approach starts with explaining the functioning of biometric systems in general terms for non-specialists. It continues with a description of the legal nature of biometric data and makes a comparison with DNA and biological material and the regulation thereof. After describing the risks, the work further reviews the opinions of data protection authorities in relation to biometric systems and current and future (EU) law. A detailed legal comparative analysis is made of the situation in Belgium, France and the Netherlands. The author concludes with an evaluation of the proportionality principle and the application of data protection law to biometric data processing operations, mainly in the private sector. Pleading for more safeguards in legislation, the author makes several suggestions for a regulatory framework aiming at reducing the risks of biometric systems. They include limitations to the collection and storage of biometric data as well as technical measures, which could influence the proportionality of the processing. The text is supported by several figures and tables providing a summary of particular points of the discussion. The book also uses the 2012 biometric vocabulary adopted by ISO and contains an extensive bibliography and literature sources.

Encyclopedia of Big Data

Data and Information in Online Environments

Eddy Covariance

Proceedings of ICABDE 2021

Data Preprocessing in Data Mining

Festschrift in Honour of Vera Pawlowsky-Glahn

Methods and Applications

This book offers practical guidelines on creating value from the application of data science based on selected artificial intelligence methods. In Part I, the author introduces a problem-driven approach to implementing AI-based data science and offers practical explanations of key technologies: machine learning, deep learning, decision trees and random forests, evolutionary computation, swarm intelligence, and intelligent agents. In Part II, he describes the main steps in creating AI-based data science solutions for business problems, including problem knowledge acquisition, data preparation, data analysis, model development, and model deployment lifecycle. Finally, in Part III the author illustrates the power of AI-based data science with successful applications in manufacturing and business. He also shows how to introduce this technology in a business setting and guides the reader on how to build the appropriate infrastructure and develop the required skillsets. The book is ideal for data scientists who will implement the proposed methodology and techniques in their projects. It is also intended to help business leaders and entrepreneurs who want to create competitive advantage by using AI-based data science, as well as academics and students looking for an industrial view of this discipline.

This is the second edition of Wil van der Aalst's seminal book on process mining, which now discusses the field also in the broader context of data science and big data approaches. It includes several additions and updates, e.g. on inductive mining techniques, the notion of alignments, a considerably expanded section on software tools and a completely new chapter of process mining in the large. It is self-contained, while at the same time covering the entire process-mining spectrum from process discovery to predictive analytics. After a general introduction to data science and process mining in Part I, Part II provides the basics of business process modeling and data mining necessary to understand the remainder of the book. Next, Part III focuses on process discovery as the most important process mining task, while Part IV moves beyond discovering the control flow of processes, highlighting conformance checking, and organizational and time perspectives. Part V offers a guide to successfully applying process mining in practice, including an introduction to the widely used open-source tool ProM and several commercial products. Lastly, Part VI takes a step back, reflecting on the material presented and the key open challenges. Overall, this book provides a comprehensive overview of the state of the art in process mining. It is intended for business process analysts, business consultants, process managers, graduate students, and BPM researchers.

The book presents studies related to artificial intelligence (AI) and its applications to process and analyze data and big data to create machines or software that can better understand business behavior, industry activities, and human health. The studies were presented at "The 2021 International Conference on Artificial Intelligence and Big Data in Digital Era" (ICABDE 2021), which was held in Ho Chi Minh City, Vietnam, during December 18-19, 2021. The studies are pointing toward the famous slogan in technology "Make everything smarter," i.e., creating machines that can understand and can communicate with humans, and they must act like humans in different aspects such as vision, communication, thinking, feeling, and acting. "A computer would deserve to be called intelligent if it could deceive a human into believing that it was human" —Alan Turing

This textbook offers an easy-to-follow, practical guide to modern data analysis using the programming language R. The chapters cover topics such as the fundamentals of programming in R, data collection and preprocessing, including web scraping, data visualization, and statistical methods, including multivariate analysis, and feature exercises at the end of each section. The text requires only basic statistics skills, as it strikes a balance between statistical and mathematical understanding and implementation in R, with a special emphasis on reproducible examples and real-world applications. This textbook is primarily intended for undergraduate students of mathematics, statistics, physics, economics, finance and business who are pursuing a career in data analytics. It will be equally valuable for master students of data science and industry professionals who want to conduct data analyses.

Big Data Analytics

Data Processing Techniques and Applications for Cyber-Physical Systems (DPTA 2019)

Elegant Graphics for Data Analysis

Optical Data Processing

With Worked Examples in R

Privacy Technologies and Policy

This book gathers a collection of high-quality peer-reviewed research papers presented at the 3rd International Conference on Data and Information Sciences (ICDIS 2021), held at Raja Balwant Singh Engineering Technical Campus, Agra, India, on May 14 – 15, 2021. In chapters written by leading researchers, developers, and practitioners from academia and industry, it covers virtually all aspects of computational sciences and information security, including central topics like artificial intelligence, cloud computing, and big data. Highlighting the latest developments and technical solutions, it will show readers from the computer industry how to capitalize on key advances in next-generation computer and communication technology.

This open access book is part of the LAMBDA Project (Learning, Applying, Multiplying Big Data Analytics), funded by the European Union, GA No. 809965. Data Analytics involves applying algorithmic processes to derive insights. Nowadays it is used in many industries to allow organizations and companies to make better decisions as well as to verify or disprove existing theories or models. The term data analytics is often used interchangeably with intelligence, statistics, reasoning, data mining, knowledge discovery, and others. The goal of this book is to introduce some of the definitions, methods, tools, frameworks, and solutions for big data processing, starting from the process of information extraction and knowledge representation, via knowledge processing and analytics to visualization, sense-making, and practical applications. Each chapter in this book addresses some pertinent aspect of the data processing chain, with a specific focus on understanding Enterprise Knowledge Graphs, Semantic Big Data Architectures, and Smart Data Analytics solutions. This book is addressed to graduate students from technical disciplines, to professional audiences following continuous education short courses, and to researchers from diverse areas following self-study courses. Basic skills in computer science, mathematics, and statistics are required.

This book is a comprehensive introduction to the methods and algorithms of modern data analytics. It provides a sound mathematical basis, discusses advantages and drawbacks of different approaches, and enables the reader to design and implement data analytics solutions for real-world applications. This book has been used for more than ten years in the Data Mining course at the Technical University of Munich. Much of the content is based on the results of industrial research and development projects at Siemens.

Artificial Intelligence in Data and Big Data ProcessingProceedings of ICABDE 2021Springer

A Survey

A Practical Guide to Measurement and Data Analysis

Applied Compositional Data Analysis

Data and Text Processing for Health and Life Sciences

An Organic Approach

Data Analysis

Intelligent Data Analysis

Continual improvements in data collection and processing have had a huge impact on brain research, producing data sets that are often large and complicated. By emphasizing a few fundamental principles, and a handful of ubiquitous techniques, Analysis of Neural Data provides a unified treatment of analytical methods that have become essential for contemporary researchers. Throughout the book ideas are illustrated with more than 100 examples drawn from the literature, ranging from electrophysiology, to neuroimaging, to behavior. By demonstrating the commonality among various statistical approaches the authors provide the crucial tools for gaining knowledge from diverse types of data. Aimed at experimentalists with only high-school level mathematics, as well as computationally-oriented neuroscientists who have limited familiarity with statistics, Analysis of Neural Data serves as both a self-contained introduction and a reference work.

This book presents modern methods and real-world applications of compositional data analysis. It covers a wide variety of topics, ranging from an updated presentation of basic concepts and ideas in compositional data analysis to recent advances in the context of complex data structures. Further, it illustrates real-world applications in numerous scientific disciplines and includes references to the latest software solutions available for compositional data analysis, thus providing a valuable and up-to-date guide for researchers and practitioners working with compositional data. Featuring selected contributions by leading experts in the field, the book is dedicated to Vera Pawlowsky-Glahn on the occasion of her 70th birthday.

This textbook for courses on function data analysis and shape data analysis describes how to define, compare, and mathematically represent shapes, with a focus on statistical modeling and inference. It is aimed at graduate students in analysis in statistics, engineering, applied mathematics, neuroscience, biology, bioinformatics, and other related areas. The interdisciplinary nature of the broad range of ideas covered—from introductory theory to algorithmic implementations and some statistical case studies—is meant to familiarize graduate students with an array of tools that are relevant in developing computational solutions for shape and related analyses.

These tools, gleaned from geometry, algebra, statistics, and computational science, are traditionally scattered across different courses, departments, and disciplines; Functional and Shape Data Analysis offers a unified, comprehensive solution by integrating the registration problem into shape analysis, better preparing graduate students for handling future scientific challenges. Recently, a data-driven and application-oriented focus on shape analysis has been trending. This text offers a self-contained treatment of this new generation of methods in shape analysis of curves. Its main focus is shape analysis of functions and curves—in one, two, and higher dimensions—both closed and open. It develops elegant Riemannian frameworks that provide both quantification of shape differences and registration of curves at the same time. Additionally, these methods are used for statistically summarizing given curve data, performing dimension reduction, and modeling observed variability. It is recommended that the reader have a background in calculus, linear algebra, numerical analysis, and computation.

Provides both rich theory and powerful applications Figures are accompanied by code required to produce them Full color figures

Approximation Theory and Algorithms for Data Analysis

Select Proceedings of ICSDP 2019

Data Science in Action

Statistical Data Analysis and Entropy

Process Mining

Statistical and Computational Methods for Scientists and Engineers

Proceedings of ICDIS 2021

This book presents the statistical analysis of compositional data using the log-ratio approach. It includes a wide range of classical and robust statistical methods adapted for compositional data analysis, such as supervised and unsupervised methods like PCA, correlation analysis, classification and regression. In addition, it considers special data structures like high-dimensional compositions and compositional tables. The methodology introduced is also frequently compared to methods which ignore the specific nature of compositional data. It focuses on practical aspects of compositional data analysis rather than on detailed theoretical derivations, thus issues like graphical visualization and preprocessing (treatment of missing values, zeros, outliers and similar artifacts) form an important part of the book. Since it is primarily intended for researchers and students from applied fields like geochemistry, chemometrics, biology and natural sciences, economics, and social sciences, all the proposed methods are accompanied by worked-out examples in R using the package robCompositions.

This book offers postgraduate and early career researchers in accounting and information systems a guide to choosing, executing and reporting appropriate data analysis methods to answer their research questions. It provides readers with a basic understanding of the steps that each method involves, and of the facets of the analysis that require special attention. Rather than presenting an exhaustive overview of the methods or explaining them in detail, the book serves as a starting point for developing data analysis skills: it provides hands-on guidelines for conducting the most common analyses and reporting results, and includes pointers to more extensive resources. Comprehensive yet succinct, the book is brief and written in a language that everyone can understand - from students to those employed by organizations wanting to study the context in which they work. It also serves as a refresher for researchers who have learned data analysis techniques previously but who need a reminder for the specific study they are involved in.

Data Preprocessing for Data Mining addresses one of the most important issues within the well-known Knowledge Discovery from Data process. Data directly taken from the source will likely have inconsistencies, errors or most importantly, it is not ready to be considered for a data mining process. Furthermore, the increasing amount of data in recent science, industry and business applications, calls to the requirement of more complex tools to analyze it. Thanks to data preprocessing, it is possible to convert the impossible into possible, adapting the data to fulfill the input demands of each data mining algorithm. Data preprocessing includes the data reduction techniques, which aim at reducing the complexity of the data, detecting or removing irrelevant and noisy elements from the data. This book is intended to review the tasks that fill the gap between the data acquisition from the source and the data mining process. A comprehensive look from a practical point of view, including basic concepts and surveying the techniques proposed in the specialized literature, is given.Each chapter is a stand-alone guide to a particular data preprocessing topic, from basic concepts and detailed descriptions of classical algorithms, to an incursion of an exhaustive catalog of recent developments. The in-depth technical descriptions make this book suitable for technical professionals, researchers, senior undergraduate and graduate students in data science, computer science and engineering. This presentation of statistical methods features extensive use of graphical displays for exploring data and for displaying the analysis. The authors demonstrate how to analyze data—showing code, graphics, and accompanying computer listings. They emphasize how to construct and interpret graphs, discuss principles of graphical design, and show how tabular results are used to confirm the visual impressions derived from the graphs. Many of the graphical formats are novel and appear here for the first time in print.

Applying Data Science

Big Data 2.0 Processing Systems

A Primer

Perception-Based Data Processing in Acoustics

The Abel Symposium 2018

Data Analytics

Statistics for Health Data Science

This book provides readers the "big picture" and a comprehensive survey of the domain of big data processing systems. For the past decade, the Hadoop framework has dominated the world of big data processing, yet recently academia and industry have started to recognize its limitations in several application domains and big data processing scenarios such as the large-scale processing of structured data, graph data and streaming data. Thus, it is now gradually being replaced by a collection of engines that are dedicated to specific verticals (e.g. structured data, graph data, and streaming data). The book explores this new wave of systems, which it refers to as Big Data 2.0 processing systems. After Chapter 1 presents the general background of the big data phenomena, Chapter 2 provides an overview of various general-purpose big data processing systems that allow their users to develop various big data processing jobs for different application domains. In turn, Chapter 3 examines various systems that have been introduced to support the SQL flavor on top of the Hadoop infrastructure and provide competing and scalable performance in the processing of large-scale structured data. Chapter 4 discusses several systems that have been designed to tackle the problem of large-scale graph processing, while the main focus of Chapter 5 is on several systems that have been designed to provide scalable solutions for processing big data streams, and on other sets of systems that have been introduced to support the development of data pipelines between various types of big data processing jobs and systems. Lastly, Chapter 6 shares conclusions and an outlook on future research challenges. Overall, the book offers a valuable reference guide for students, researchers and professionals in the domain of big data processing systems. Further, its comprehensive content will hopefully encourage readers to pursue further research on the subject.

This encyclopedia will be an essential resource for our times, reflecting the fact that we currently are living in an expanding data-driven world. Technological advancements and other related trends are contributing to the production of an astoundingly large and exponentially increasing collection of data and information, referred to in popular vernacular as "Big Data." Social media and crowdsourcing platforms and various applications — "apps" — are producing reams of information from the instantaneous transactions and input of millions and millions of people around the globe. The Internet-of-Things (IoT), which is expected to comprise tens of billions of objects by the end of this decade, is actively sensing real-time intelligence on nearly every aspect of our lives and environment. The Global Positioning System (GPS) and other location-aware technologies are producing data that is specific down to particular latitude and longitude coordinates and seconds of the day. Large-scale instruments, such as the Large Hadron Collider (LHC), are collecting massive amounts of data on our planet and even distant corners of the visible universe. Digitization is being used to convert large collections of documents from print to digital format, giving rise to large archives of unstructured data. Innovations in technology, in the areas of Cloud and molecular computing, Artificial Intelligence/Machine Learning, and Natural Language Processing (NLP), to name only a few, also are greatly expanding our capacity to store, manage, and process Big Data. In this context, the Encyclopedia of Big Data is being offered in recognition of a world that is rapidly moving from gigabytes to terabytes to petabytes and beyond. While indeed large data sets have long been around and in use in a variety of fields, the era of Big Data in which we now live departs from the past in a number of key respects and with this departure comes a fresh set of challenges and opportunities that cut across and affect multiple sectors and disciplines, and the public at large. With expanded analytical capacities at hand, Big Data is now being used for scientific inquiry and experimentation in nearly every (if not all) disciplines, from the social sciences to the humanities to the natural sciences, and more. Moreover, the use of Big Data has been well established beyond the Ivory Tower. In today's economy, businesses simply cannot be competitive without engaging Big Data in one way or another in support of operations, management, planning, or simply basic hiring decisions. In all levels of government, Big Data is being used to engage citizens and to guide policy making in pursuit of the interests of the public and society in general. Moreover, the changing nature of Big Data also raises new issues and concerns related to, for example, privacy, liability, security, access, and even the veracity of the data itself. Given the complex issues attending Big Data, there is a real need for a reference book that covers the subject from a multi-disciplinary, cross-sectoral, comprehensive, and international perspective. The Encyclopedia of Big Data will address this need and will be the first of such reference books to do so. Featuring some 500 entries, from "Access" to "Zillow," the Encyclopedia will serve as a fundamental resource for researchers and students, for decision makers and leaders, and for business analysts and purveyors. Developed for those in academia, industry, and government, and others with a general interest in Big Data, the encyclopedia will be aimed especially at those involved in its collection, analysis, and use. Ultimately, the Encyclopedia of Big Data will provide a common platform and language covering the breadth and depth of the topic for different segments, sectors, and disciplines.

This book gathers the proceedings of the 2018 Abel Symposium, which was held in Geiranger, Norway, on June 4-8, 2018. The symposium offered an overview of the emerging field of "Topological Data Analysis". This volume presents papers on various research directions, notably including applications in neuroscience, materials science, cancer biology, and immune response. Providing an essential snapshot of the status quo, it represents a valuable asset for practitioners and those considering entering the field.

This monograph provides novel insights into cognitive mechanisms underlying the processing of sound and music in different environments. A solid understanding of these mechanisms is vital for numerous technological applications such as for example information retrieval from distributed musical databases or building expert systems. In order to investigate the cognitive mechanisms of music perception fundamentals of hearing psychophysiology and principles of music perception are presented. In addition, some computational intelligence methods are reviewed, such as rough sets, fuzzy logic, artificial neural networks, decision trees and genetic algorithms. The applications of hybrid decision systems to problem solving in music and acoustics are exemplified and discussed on the basis of obtained experimental results.

Mathematical Foundations for Data Analysis

Intelligent Data Analysis for COVID-19 Pandemic

Advances in Compositional Data Analysis