

Statistics For High Dimensional Data Methods Theory And Applications

Yeah, reviewing a ebook **Statistics For High Dimensional Data Methods Theory And Applications** could grow your near associates listings. This is just one of the solutions for you to be successful. As understood, success does not suggest that you have extraordinary points.

Comprehending as with ease as pact even more than additional will have the funds for each success. next-door to, the proclamation as capably as keenness of this Statistics For High Dimensional Data Methods Theory And Applications can be taken as competently as picked to act.

An Introduction to Statistical Learning Gareth James 2013-06-24 An Introduction to Statistical Learning provides an accessible overview of the field of statistical learning, an essential toolset for making sense of the vast and complex data sets that have emerged in fields ranging from biology to finance to marketing to astrophysics in the past twenty years. This book presents some of the most important modeling and prediction techniques, along with relevant applications. Topics include linear regression, classification, resampling methods, shrinkage approaches, tree-based methods, support vector machines, clustering, and more. Color graphics and real-world examples are used to illustrate the methods presented. Since the goal of this textbook is to facilitate the use of these statistical learning techniques by practitioners in science, industry, and other fields, each chapter contains a tutorial on implementing the analyses and methods presented in R, an extremely popular open source statistical software platform. Two of the authors co-wrote *The Elements of Statistical Learning* (Hastie, Tibshirani and Friedman, 2nd edition 2009), a popular reference book for statistics and machine learning researchers. An Introduction to Statistical Learning covers many of the same topics, but at a level accessible to a much broader audience. This book is targeted at statisticians and non-statisticians alike who wish to use cutting-edge statistical learning techniques to analyze their data. The text assumes only a previous course in linear regression and no knowledge of matrix algebra.

Functional and High-Dimensional Statistics and Related Fields Germán Aneiros 2020-06-19 This book presents the latest research on the statistical analysis of functional, high-dimensional and other complex data, addressing methodological and computational aspects, as well as real-world applications. It covers topics like classification, confidence bands, density estimation, depth, diagnostic tests, dimension reduction, estimation on manifolds, high- and infinite-dimensional statistics, inference on functional data, networks, operatorial statistics, prediction, regression, robustness, sequential learning, small-ball probability, smoothing, spatial data, testing, and topological object data analysis, and includes applications in automobile engineering, criminology, drawing recognition, economics, environmetrics, medicine, mobile phone data, spectrometrics and urban environments. The book gathers selected, refereed contributions presented at the Fifth International Workshop on Functional and Operatorial Statistics (IWFOS) in Brno, Czech Republic. The workshop was originally to be held on June 24-26, 2020, but had to be postponed as a consequence of the COVID-19 pandemic. Initiated by the Working Group on Functional and Operatorial Statistics at the University of Toulouse in 2008, the IWFOS workshops provide a forum to discuss the latest trends and advances in functional statistics and related fields, and foster the exchange of ideas and international collaboration in the field.

Sparse Modeling for Image and Vision Processing Julien Mairal 2014-12-19 Sparse Modeling for Image and Vision Processing offers a self-contained view of sparse modeling for visual recognition and image processing. More specifically, it focuses on applications where the dictionary is learned and adapted to data, yielding a compact representation that has been successful in various contexts.

Principles and Theory for Data Mining and Machine Learning Bertrand Clarke 2009-07-21 Extensive treatment of the most up-to-date topics Provides the theory and concepts behind popular and emerging methods Range of topics drawn from Statistics, Computer Science, and Electrical Engineering

High-dimensional Data Analysis Tianwen Tony Cai 2011 Over the last few years, significant developments have been taking place in high-dimensional data analysis, driven primarily by a wide range of applications in many fields such as genomics and signal processing. In particular, substantial advances have been made in the areas of feature selection, covariance estimation, classification and regression. This book intends to examine important issues arising from high-dimensional data analysis to explore key ideas for statistical inference and prediction. It is structured around topics on multiple hypothesis testing, feature selection, regression, classification, dimension reduction, as well as applications in survival analysis and biomedical research. The book will appeal to graduate students and new researchers interested in the plethora of opportunities available in high-dimensional data analysis.

Big and Complex Data Analysis S. Ejaz Ahmed 2017-03-21 This volume conveys some of the surprises, puzzles and success stories in high-dimensional and complex data analysis and related fields. Its peer-reviewed contributions showcase recent advances in variable selection, estimation and prediction strategies for a host of useful models, as well as essential new developments in the field. The continued and rapid advancement of modern technology now allows scientists to collect data of increasingly unprecedented size and complexity. Examples include epigenomic data, genomic data, proteomic data, high-resolution image data, high-frequency financial data, functional and longitudinal data, and network data. Simultaneous variable selection and estimation is one of the key statistical problems involved in analyzing such big and complex data. The purpose of this book is to stimulate research and foster interaction between researchers in the area of high-dimensional data analysis. More concretely, its goals are to: 1) highlight and expand the breadth of existing methods in big data and high-dimensional data analysis and their potential for the advancement of both the mathematical and statistical sciences; 2) identify important directions for future research in the theory of regularization methods, in algorithmic development, and in methodologies for different application areas; and 3) facilitate collaboration between theoretical and subject-specific researchers.

Statistical Inference from High Dimensional Data Carlos Fernandez-Lozano 2021-04-28 • Real-world problems can be high-dimensional, complex, and noisy • More data does not imply more information • Different approaches deal with the so-called curse of dimensionality to reduce irrelevant information • A process with multidimensional information is not necessarily easy to interpret nor process • In some real-world applications, the number of elements of a class is clearly lower than the other. The models tend to assume that the importance of the analysis belongs to the majority class and this is not usually the truth • The analysis of complex diseases such as cancer are focused on more-than-one dimensional omic data • The increasing amount of data thanks to the reduction of cost of the high-throughput experiments opens up a new era for integrative data-driven approaches • Entropy-based approaches are of interest to

reduce the dimensionality of high-dimensional data

Multivariate Statistics Yasunori Fujikoshi 2011-08-15 A comprehensive examination of high-dimensional analysis of multivariate methods and their real-world applications Multivariate Statistics: High-Dimensional and Large-Sample Approximations is the first book of its kind to explore how classical multivariate methods can be revised and used in place of conventional statistical tools. Written by prominent researchers in the field, the book focuses on high-dimensional and large-scale approximations and details the many basic multivariate methods used to achieve high levels of accuracy. The authors begin with a fundamental presentation of the basic tools and exact distributional results of multivariate statistics, and, in addition, the derivations of most distributional results are provided. Statistical methods for high-dimensional data, such as curve data, spectra, images, and DNA microarrays, are discussed. Bootstrap approximations from a methodological point of view, theoretical accuracies in MANOVA tests, and model selection criteria are also presented. Subsequent chapters feature additional topical coverage including: High-dimensional approximations of various statistics High-dimensional statistical methods Approximations with computable error bound Selection of variables based on model selection approach Statistics with error bounds and their appearance in discriminant analysis, growth curve models, generalized linear models, profile analysis, and multiple comparison Each chapter provides real-world applications and thorough analyses of the real data. In addition, approximation formulas found throughout the book are a useful tool for both practical and theoretical statisticians, and basic results on exact distributions in multivariate analysis are included in a comprehensive, yet accessible, format. Multivariate Statistics is an excellent book for courses on probability theory in statistics at the graduate level. It is also an essential reference for both practical and theoretical statisticians who are interested in multivariate analysis and who would benefit from learning the applications of analytical probabilistic methods in statistics.

Geometric Structure of High-Dimensional Data and Dimensionality Reduction Jianzhong Wang 2012-04-28 "Geometric Structure of High-Dimensional Data and Dimensionality Reduction" adopts data geometry as a framework to address various methods of dimensionality reduction. In addition to the introduction to well-known linear methods, the book moreover stresses the recently developed nonlinear methods and introduces the applications of dimensionality reduction in many areas, such as face recognition, image segmentation, data classification, data visualization, and hyperspectral imagery data analysis. Numerous tables and graphs are included to illustrate the ideas, effects, and shortcomings of the methods. MATLAB code of all dimensionality reduction algorithms is provided to aid the readers with the implementations on computers. The book will be useful for mathematicians, statisticians, computer scientists, and data analysts. It is also a valuable handbook for other practitioners who have a basic background in mathematics, statistics and/or computer algorithms, like internet search engine designers, physicists, geologists, electronic engineers, and economists. Jianzhong Wang is a Professor of Mathematics at Sam Houston State University, U.S.A.

High-Dimensional Statistics Martin J. Wainwright 2019-02-21 A coherent introductory text from a groundbreaking researcher, focusing on clarity and motivation to build intuition and understanding.

Multivariate Statistical Machine Learning Methods for Genomic Prediction Osval Antonio Montesinos López 2022-02-14 This book is open access under a CC BY 4.0 license This open access book brings together the latest genome base prediction models currently being used by statisticians, breeders and data scientists. It provides an accessible way to understand the theory behind each statistical learning tool, the required pre-processing, the basics of model building, how to train statistical learning methods, the basic R scripts needed to implement each statistical learning tool, and the output of each tool. To do so, for each tool the book provides background theory, some elements of the R statistical software for its implementation, the conceptual underpinnings, and at least two illustrative examples with data from real-world genomic selection experiments. Lastly, worked-out examples help readers check their own comprehension. The book will greatly appeal to readers in plant (and animal) breeding, geneticists and statisticians, as it provides in a very accessible way the necessary theory, the appropriate R code, and illustrative examples for a complete understanding of each statistical learning tool. In addition, it weighs the advantages and disadvantages of each tool. **Geometric and Topological Inference** Jean-Daniel Boissonnat 2018-09-27 A rigorous introduction to geometric and topological inference, for anyone interested in a geometric approach to data science.

Introduction to High-Dimensional Statistics Christophe Giraud 2014-12-17 Ever-greater computing technologies have given rise to an exponentially growing volume of data. Today massive data sets (with potentially thousands of variables) play an important role in almost every branch of modern human activity, including networks, finance, and genetics. However, analyzing such data has presented a challenge for statisticians and data analysts and has required the development of new statistical methods capable of separating the signal from the noise. Introduction to High-Dimensional Statistics is a concise guide to state-of-the-art models, techniques, and approaches for handling high-dimensional data. The book is intended to expose the reader to the key concepts and ideas in the most simple settings possible while avoiding unnecessary technicalities. Offering a succinct presentation of the mathematical foundations of high-dimensional statistics, this highly accessible text: Describes the challenges related to the analysis of high-dimensional data Covers cutting-edge statistical methods including model selection, sparsity and the lasso, aggregation, and learning theory Provides detailed exercises at the end of every chapter with collaborative solutions on a wikisite Illustrates concepts with simple but clear practical examples Introduction to High-Dimensional Statistics is suitable for graduate students and researchers interested in discovering modern statistics for massive data. It can be used as a graduate text or for self-study.

Bayesian and High-Dimensional Global Optimization Anatoly Zhigljavsky 2021-03-02 Accessible to a variety of readers, this book is of interest to specialists, graduate students and researchers in mathematics, optimization, computer science, operations research, management science, engineering and other applied areas interested in solving optimization problems. Basic principles, potential and boundaries of applicability of stochastic global optimization

techniques are examined in this book. A variety of issues that face specialists in global optimization are explored, such as multidimensional spaces which are frequently ignored by researchers. The importance of precise interpretation of the mathematical results in assessments of optimization methods is demonstrated through examples of convergence in probability of random search. Methodological issues concerning construction and applicability of stochastic global optimization methods are discussed, including the one-step optimal average improvement method based on a statistical model of the objective function. A significant portion of this book is devoted to an analysis of high-dimensional global optimization problems and the so-called 'curse of dimensionality'. An examination of the three different classes of high-dimensional optimization problems, the geometry of high-dimensional balls and cubes, very slow convergence of global random search algorithms in large-dimensional problems, and poor uniformity of the uniformly distributed sequences of points are included in this book.

Computer Age Statistical Inference, Student Edition Bradley Efron 2021-06-17 The twenty-first century has seen a breathtaking expansion of statistical methodology, both in scope and influence. 'Data science' and 'machine learning' have become familiar terms in the news, as statistical methods are brought to bear upon the enormous data sets of modern science and commerce. How did we get here? And where are we going? How does it all fit together? Now in paperback and fortified with exercises, this book delivers a concentrated course in modern statistical thinking. Beginning with classical inferential theories - Bayesian, frequentist, Fisherian - individual chapters take up a series of influential topics: survival analysis, logistic regression, empirical Bayes, the jackknife and bootstrap, random forests, neural networks, Markov Chain Monte Carlo, inference after model selection, and dozens more. The distinctly modern approach integrates methodology and algorithms with statistical inference. Each chapter ends with class-tested exercises, and the book concludes with speculation on the future direction of statistics and data science.

Statistical Foundations of Data Science Jianqing Fan 2020-09-21 Statistical Foundations of Data Science gives a thorough introduction to commonly used statistical models, contemporary statistical machine learning techniques and algorithms, along with their mathematical insights and statistical theories. It aims to serve as a graduate-level textbook and a research monograph on high-dimensional statistics, sparsity and covariance learning, machine learning, and statistical inference. It includes ample exercises that involve both theoretical studies as well as empirical applications. The book begins with an introduction to the stylized features of big data and their impacts on statistical analysis. It then introduces multiple linear regression and expands the techniques of model building via nonparametric regression and kernel tricks. It provides a comprehensive account on sparsity explorations and model selections for multiple regression, generalized linear models, quantile regression, robust regression, hazards regression, among others. High-dimensional inference is also thoroughly addressed and so is feature screening. The book also provides a comprehensive account on high-dimensional covariance estimation, learning latent factors and hidden structures, as well as their applications to statistical estimation, inference, prediction and machine learning problems. It also introduces thoroughly statistical machine learning theory and methods for classification, clustering, and prediction. These include CART, random forests, boosting, support vector machines, clustering algorithms, sparse PCA, and deep learning.

The Concentration of Measure Phenomenon Michel Ledoux 2005 It was undoubtedly a necessary task to collect all the results on the concentration of measure during the past years in a monograph. The author did this very successfully and the book is an important contribution to the topic. It will surely influence further research in this area considerably. The book is very well written, and it was a great pleasure for the reviewer to read it. --Mathematical Reviews The observation of the concentration of measure phenomenon is inspired by isoperimetric inequalities. A familiar example is the way the uniform measure on the standard sphere S^n becomes concentrated around the equator as the dimension gets large. This property may be interpreted in terms of functions on the sphere with small oscillations, an idea going back to Levy. The phenomenon also occurs in probability, as a version of the law of large numbers, due to Emile Borel. This book offers the basic techniques and examples of the concentration of measure phenomenon. The concentration of measure phenomenon was put forward in the early seventies by V. Milman in the asymptotic geometry of Banach spaces. It is of powerful interest in applications in various areas, such as geometry, functional analysis and infinite-dimensional integration, discrete mathematics and complexity theory, and probability theory. Particular emphasis is on geometric, functional, and probabilistic tools to reach and describe measure concentration in a number of settings. The book presents concentration functions and inequalities, isoperimetric and functional examples, spectrum and topological applications, product measures, entropic and transportation methods, as well as aspects of M. Talagrand's deep investigation of concentration in product spaces and its application in discrete mathematics and probability theory, supremum of Gaussian and empirical processes, spin glass, random matrices, etc. Prerequisites are a basic background in measure theory, functional analysis, and probability theory.

Statistical Learning with Sparsity Trevor Hastie 2015-05-07 Discover New Methods for Dealing with High-Dimensional Data A sparse statistical model has only a small number of nonzero parameters or weights; therefore, it is much easier to estimate and interpret than a dense model. Statistical Learning with Sparsity: The Lasso and Generalizations presents methods that exploit sparsity to help recover the underlying signal in a set of data. Top experts in this rapidly evolving field, the authors describe the lasso for linear regression and a simple coordinate descent algorithm for its computation. They discuss the application of l1 penalties to generalized linear models and support vector machines, cover generalized penalties such as the elastic net and group lasso, and review numerical methods for optimization. They also present statistical inference methods for fitted (lasso) models, including the bootstrap, Bayesian methods, and recently developed approaches. In addition, the book examines matrix decomposition, sparse multivariate analysis, graphical models, and compressed sensing. It concludes with a survey of theoretical results for the lasso. In this age of big data, the number of features measured on a person or object can be large and might be larger than the number of observations. This book shows how the sparsity assumption allows us to tackle these problems and extract useful and reproducible patterns from big datasets. Data analysts, computer scientists, and theorists will appreciate this thorough and up-to-date treatment of sparse statistical modeling.

Large Sample Covariance Matrices and High-Dimensional Data Analysis Jianfeng Yao 2015-03-26 High-dimensional data appear in many fields, and their analysis has become increasingly important in modern statistics. However, it has long been observed that several well-known methods in multivariate analysis become inefficient, or even misleading, when the data dimension p is larger than, say, several tens. A seminal example is the well-known inefficiency of Hotelling's T^2 -test in such cases. This example shows that classical large sample limits may no longer hold for high-dimensional data; statisticians must seek new limiting theorems in these instances. Thus, the theory of random matrices (RMT) serves as a much-needed and welcome alternative framework. Based on the authors' own research, this book provides a first-hand introduction to new high-dimensional statistical methods derived from RMT. The book begins with a detailed introduction

to useful tools from RMT, and then presents a series of high-dimensional problems with solutions provided by RMT methods.

Introduction to High-Dimensional Statistics Christophe Giraud 2021-08-26 Praise for the first edition: "[This book] succeeds singularly at providing a structured introduction to this active field of research. ... it is arguably the most accessible overview yet published of the mathematical ideas and principles that one needs to master to enter the field of high-dimensional statistics. ... recommended to anyone interested in the main results of current research in high-dimensional statistics as well as anyone interested in acquiring the core mathematical skills to enter this area of research." --Journal of the American Statistical Association Introduction to High-Dimensional Statistics, Second Edition preserves the philosophy of the first edition: to be a concise guide for students and researchers discovering the area and interested in the mathematics involved. The main concepts and ideas are presented in simple settings, avoiding thereby unessential technicalities. High-dimensional statistics is a fast-evolving field, and much progress has been made on a large variety of topics, providing new insights and methods. Offering a succinct presentation of the mathematical foundations of high-dimensional statistics, this new edition: Offers revised chapters from the previous edition, with the inclusion of many additional materials on some important topics, including compress sensing, estimation with convex constraints, the slope estimator, simultaneously low-rank and row-sparse linear regression, or aggregation of a continuous set of estimators. Introduces three new chapters on iterative algorithms, clustering, and minimax lower bounds. Provides enhanced appendices, minimax lower-bounds mainly with the addition of the Davis-Kahan perturbation bound and of two simple versions of the Hanson-Wright concentration inequality. Covers cutting-edge statistical methods including model selection, sparsity and the Lasso, iterative hard thresholding, aggregation, support vector machines, and learning theory. Provides detailed exercises at the end of every chapter with collaborative solutions on a wiki site. Illustrates concepts with simple but clear practical examples.

High-Dimensional Covariance Estimation Mohsen Pourahmadi 2013-05-28 Methods for estimating sparse and large covariance matrices Covariance and correlation matrices play fundamental roles in every aspect of the analysis of multivariate data collected from a variety of fields including business and economics, health care, engineering, and environmental and physical sciences. High-Dimensional Covariance Estimation provides accessible and comprehensive coverage of the classical and modern approaches for estimating covariance matrices as well as their applications to the rapidly developing areas lying at the intersection of statistics and machine learning. Recently, the classical sample covariance methodologies have been modified and improved upon to meet the needs of statisticians and researchers dealing with large correlated datasets. High-Dimensional Covariance Estimation focuses on the methodologies based on shrinkage, thresholding, and penalized likelihood with applications to Gaussian graphical models, prediction, and mean-variance portfolio management. The book relies heavily on regression-based ideas and interpretations to connect and unify many existing methods and algorithms for the task. High-Dimensional Covariance Estimation features chapters on: Data, Sparsity, and Regularization Regularizing the Eigenstructure Banding, Tapering, and Thresholding Covariance Matrices Sparse Gaussian Graphical Models Multivariate Regression The book is an ideal resource for researchers in statistics, mathematics, business and economics, computer sciences, and engineering, as well as a useful text or supplement for graduate-level courses in multivariate analysis, covariance estimation, statistical learning, and high-dimensional data analysis.

High-Dimensional Covariance Matrix Estimation Aygul Zagidullina

Foundations of Data Science Avrim Blum 2020-01-23 This book provides an introduction to the mathematical and algorithmic foundations of data science, including machine learning, high-dimensional geometry, and analysis of large networks. Topics include the counterintuitive nature of data in high dimensions, important linear algebraic techniques such as singular value decomposition, the theory of random walks and Markov chains, the fundamentals of and important algorithms for machine learning, algorithms and analysis for clustering, probabilistic models for large networks, representation learning including topic modelling and non-negative matrix factorization, wavelets and compressed sensing. Important probabilistic techniques are developed including the law of large numbers, tail inequalities, analysis of random projections, generalization guarantees in machine learning, and moment methods for analysis of phase transitions in large random graphs. Additionally, important structural and complexity measures are discussed such as matrix norms and VC-dimension. This book is suitable for both undergraduate and graduate courses in the design and analysis of algorithms for data.

Principles and Methods for Data Science 2020-05-28 Principles and Methods for Data Science, Volume 43 in the Handbook of Statistics series, highlights new advances in the field, with this updated volume presenting interesting and timely topics, including Competing risks, aims and methods, Data analysis and mining of microbial community dynamics, Support Vector Machines, a robust prediction method with applications in bioinformatics, Bayesian Model Selection for Data with High Dimension, High dimensional statistical inference: theoretical development to data analytics, Big data challenges in genomics, Analysis of microarray gene expression data using information theory and stochastic algorithm, Hybrid Models, Markov Chain Monte Carlo Methods: Theory and Practice, and more. Provides the authority and expertise of leading contributors from an international board of authors Presents the latest release in the Handbook of Statistics series Updated release includes the latest information on Principles and Methods for Data Science

Fundamentals of High-Dimensional Statistics Johannes Lederer 2021-11-16 This textbook provides a step-by-step introduction to the tools and principles of high-dimensional statistics. Each chapter is complemented by numerous exercises, many of them with detailed solutions, and computer labs in R that convey valuable practical insights. The book covers the theory and practice of high-dimensional linear regression, graphical models, and inference, ensuring readers have a smooth start in the field. It also offers suggestions for further reading. Given its scope, the textbook is intended for beginning graduate and advanced undergraduate students in statistics, biostatistics, and bioinformatics, though it will be equally useful to a broader audience.

Nonlinear Dimensionality Reduction John A. Lee 2007-10-31 This book describes established and advanced methods for reducing the dimensionality of numerical databases. Each description starts from intuitive ideas, develops the necessary mathematical details, and ends by outlining the algorithmic implementation. The text provides a lucid summary of facts and concepts relating to well-known methods as well as recent developments in nonlinear dimensionality reduction. Methods are all described from a unifying point of view, which helps to highlight their respective strengths and shortcomings. The presentation will appeal to statisticians, computer scientists and data analysts, and other practitioners having a basic background in statistics or computational learning.

High-dimensional Statistics Johannes Lederer 2020 "An Introduction to Regularized Estimation in High Dimensions considers statistical theory, methods, and algorithms for large and complex data. The main focus is on regularized estimators, which are at the cusp of entering the statistical toolkits of almost all scientific disciplines. This book

provides clear expositions, motivational introductions to each chapter, rigorous step-by-step proofs, and comprehensive exercise sets with fully worked out solutions. These features make this book ideal for graduate level courses.

Moreover, the book also discusses cutting-edge topics, such as aspects of inference, robustness, and tuning parameters. The book also contains results and insights that are new altogether, including improvements of existing theories and novel connections among different methods, which are organized into special chapters for those wishing to advance their reading in the field."--

Inverse Problems and High-Dimensional Estimation Pierre Alquier 2011-06-07 The "Stats in the Château" summer school was held at the CRC château on the campus of HEC Paris, Jouy-en-Josas, France, from August 31 to September 4, 2009. This event was organized jointly by faculty members of three French academic institutions – ENSAE ParisTech, the Ecole Polytechnique ParisTech, and HEC Paris – which cooperate through a scientific foundation devoted to the decision sciences. The scientific content of the summer school was conveyed in two courses, one by Laurent Cavalier (Université Aix-Marseille I) on "Ill-posed Inverse Problems", and one by Victor Chernozhukov (Massachusetts Institute of Technology) on "High-dimensional Estimation with Applications to Economics". Ten invited researchers also presented either reviews of the state of the art in the field or of applications, or original research contributions. This volume contains the lecture notes of the two courses. Original research articles and a survey complement these lecture notes. Applications to economics are discussed in various contributions.

Statistical Diagnostics for Cancer Matthias Dehmer 2013-02-25

Introduction to High-Dimensional Statistics Christophe Giraud 2014-12-17 Ever-greater computing technologies have given rise to an exponentially growing volume of data. Today massive data sets (with potentially thousands of variables) play an important role in almost every branch of modern human activity, including networks, finance, and genetics. However, analyzing such data has presented a challenge for statisticians and data analysts and has required the development of new statistical methods capable of separating the signal from the noise. Introduction to High-Dimensional Statistics is a concise guide to state-of-the-art models, techniques, and approaches for handling high-dimensional data. The book is intended to expose the reader to the key concepts and ideas in the most simple settings possible while avoiding unnecessary technicalities. Offering a succinct presentation of the mathematical foundations of high-dimensional statistics, this highly accessible text: Describes the challenges related to the analysis of high-dimensional data Covers cutting-edge statistical methods including model selection, sparsity and the lasso, aggregation, and learning theory Provides detailed exercises at the end of every chapter with collaborative solutions on a wikisite Illustrates concepts with simple but clear practical examples Introduction to High-Dimensional Statistics is suitable for graduate students and researchers interested in discovering modern statistics for massive data. It can be used as a graduate text or for self-study.

Analysis of Multivariate and High-Dimensional Data Inge Koch 2013-12-02 'Big data' poses challenges that require both classical multivariate methods and contemporary techniques from machine learning and engineering. This modern text equips you for the new world - integrating the old and the new, fusing theory and practice and bridging the gap to statistical learning. The theoretical framework includes formal statements that set out clearly the guaranteed 'safe operating zone' for the methods and allow you to assess whether data is in the zone, or near enough. Extensive examples showcase the strengths and limitations of different methods with small classical data, data from medicine, biology, marketing and finance, high-dimensional data from bioinformatics, functional data from proteomics, and simulated data. High-dimension low-sample-size data gets special attention. Several data sets are revisited repeatedly to allow comparison of methods. Generous use of colour, algorithms, Matlab code, and problem sets complete the package. Suitable for master's/graduate students in statistics and researchers in data-rich disciplines.

High-Dimensional Probability Roman Vershynin 2018-09-27 An integrated package of powerful probabilistic tools and key applications in modern mathematical data science.

Statistical Analysis for High-Dimensional Data Arnaldo Frigessi 2016-02-16 This book features research contributions from The Abel Symposium on Statistical Analysis for High Dimensional Data, held in Nyvågar, Lofoten, Norway, in May 2014. The focus of the symposium was on statistical and machine learning methodologies specifically developed for inference in "big data" situations, with particular reference to genomic applications. The contributors, who are among the most prominent researchers on the theory of statistics for high dimensional inference, present new theories and methods, as well as challenging applications and computational solutions. Specific themes include, among others, variable selection and screening, penalised regression, sparsity, thresholding, low dimensional structures, computational challenges, non-convex situations, learning graphical models, sparse covariance and precision matrices, semi- and non-parametric formulations, multiple testing, classification, factor models, clustering, and preselection. Highlighting cutting-edge research and casting light on future research directions, the contributions will benefit graduate students and researchers in computational biology, statistics and the machine learning community.

Model-Based Clustering and Classification for Data Science Charles Bouveyron 2019-09-30 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. *Contributions in infinite-dimensional statistics and related topics* Enea G. Bongiorno 2014-05-21 The interest towards Functional and Operatorial Statistics, and, more in general, towards infinite-dimensional statistics has dramatically increased in the statistical community and in many other applied scientific areas where people faces functional data. This volume collects the works selected and presented at the Third Edition of the International Workshop on Functional and Operatorial Statistics held in Stresa, Italy, from the 19th to the 21st of June 2014 (IWFOS'2014). The meeting represents an opportunity of bringing together leading researchers active on these topics both for what concerns theoretical aspects and a wide range of applications in various fields. To promote collaborations with other important strictly related areas of infinite-dimensional Statistics, such as High Dimensional Statistics and Model Selection Procedures, this book hosts works in the latter research subjects too.

High-Dimensional Data Analysis with Low-Dimensional Models John Wright 2021-12-31 Connects fundamental mathematical theory with real-world problems, through efficient and scalable optimization algorithms.

Modern Directional Statistics Christophe Ley 2017-08-03 Modern Directional Statistics collects important advances in methodology and theory for directional statistics over the last two decades. It provides a detailed overview and analysis of recent results that can help both researchers and practitioners. Knowledge of multivariate statistics eases the reading but is not mandatory. The field of directional statistics has received a lot of attention over the past two decades, due to new demands from domains such as life sciences or machine learning, to the availability of massive data sets requiring adapted statistical techniques, and to technological advances. This book covers important progresses in distribution theory, high-dimensional statistics, kernel density estimation, efficient inference on directional supports, and computational and graphical methods. Christophe Ley is professor of mathematical statistics at Ghent University. His research interests include semi-parametrically efficient inference, flexible modeling, directional statistics and the study of asymptotic approximations via Stein's Method. His achievements include the Marie-Jeanne Laurent-Duhamel prize of the Société Française de Statistique and an elected membership at the International Statistical Institute. He is associate editor for the journals Computational Statistics & Data Analysis and Econometrics and Statistics. Thomas Verdebout is professor of mathematical statistics at Université libre de Bruxelles (ULB). His main research interests are semi-parametric statistics, high-dimensional statistics, directional statistics and rank-based procedures. He has won an annual prize of the Belgian Academy of Sciences and is an elected member of the International Statistical Institute. He is associate editor for the journals Statistics and Probability Letters and Journal of Multivariate Analysis.

Applied Biclustering Methods for Big and High-Dimensional Data Using R Adetayo Kasim 2016-10-03 Proven Methods for Big Data Analysis As big data has become standard in many application areas, challenges have arisen related to methodology and software development, including how to discover meaningful patterns in the vast amounts of data. Addressing these problems, Applied Biclustering Methods for Big and High-Dimensional Data Using R shows how to apply biclustering methods to find local patterns in a big data matrix. The book presents an overview of data analysis using biclustering methods from a practical point of view. Real case studies in drug discovery, genetics, marketing research, biology, toxicity, and sports illustrate the use of several biclustering methods. References to technical details of the methods are provided for readers who wish to investigate the full theoretical background. All the methods are accompanied with R examples that show how to conduct the analyses. The examples, software, and other materials are available on a supplementary website.

Mathematical Foundations of Infinite-Dimensional Statistical Models Evarist Giné 2021-03-25 In nonparametric and high-dimensional statistical models, the classical Gauss–Fisher–Le Cam theory of the optimality of maximum likelihood estimators and Bayesian posterior inference does not apply, and new foundations and ideas have been developed in the past several decades. This book gives a coherent account of the statistical theory in infinite-dimensional parameter spaces. The mathematical foundations include self-contained 'mini-courses' on the theory of Gaussian and empirical processes, approximation and wavelet theory, and the basic theory of function spaces. The theory of statistical inference in such models - hypothesis testing, estimation and confidence sets - is presented within the minimax paradigm of decision theory. This includes the basic theory of convolution kernel and projection estimation, but also Bayesian nonparametrics and nonparametric maximum likelihood estimation. In a final chapter the theory of adaptive inference in nonparametric models is developed, including Lepski's method, wavelet thresholding, and adaptive inference for self-similar functions. Winner of the 2017 PROSE Award for Mathematics.

Statistics for High-Dimensional Data Peter Bühlmann 2011-06-08 Modern statistics deals with large and complex data sets, and consequently with models containing a large number of parameters. This book presents a detailed account of recently developed approaches, including the Lasso and versions of it for various models, boosting methods, undirected graphical modeling, and procedures controlling false positive selections. A special characteristic of the book is that it contains comprehensive mathematical theory on high-dimensional statistics combined with methodology, algorithms and illustrations with real data examples. This in-depth approach highlights the methods' great potential and practical applicability in a variety of settings. As such, it is a valuable resource for researchers, graduate students and experts in statistics, applied mathematics and computer science.