Density Estimation For Statistics And Data Analysis Ned

This book describes computational problems related to kernel density estimation (KDE) – one of the most important and widely used data smoothing techniques. A very detailed description of novel FFT-based algorithms for both KDE computations and bandwidth selection are presented. The theory of KDE appears to have matured and is now well developed and understood. However, there is not much progress observed in terms of performance improvements. This book is an attempt to remedy this. The book primarily addresses researchers and advanced graduate or postgraduate students who are interested in KDE and its computational aspects. The book contains both some background and much more sophisticated material, hence also more experienced researchers in the KDE area may find it interesting. The presented material is richly illustrated with many numerical examples using both artificial and real datasets. Also, a number of practical applications related to KDE are presented.

Developed from lecture notes and ready to be used for a course on the graduate level, this concise text aims to introduce the fundamental concepts of nonparametric estimation theory while maintaining the exposition suitable for a first approach in the field.

Unique blend of asymptotic theory and small sample practice through simulation experiments and data analysis. Novel reproducing kernel Hilbert space methods for the analysis of smoothing splines and local polynomials. Leading to uniform error bounds and honest confidence bands for the mean function using smoothing splines Exhaustive exposition of algorithms, including the Kalman filter, for the computation of smoothing splines of arbitrary order.

This book gives a systematic, comprehensive, and unified account of modern nonparametric statistics of density estimation, nonparametric regression, filtering signals, and time series analysis. The companion software package, available over the Internet, brings all of the discussed topics into the realm of interactive research. Virtually every claim and development mentioned in the book is illustrated with graphs which are available for the reader to reproduce and modify, making the material fully transparent and allowing for complete interactivity.

Although there has been a surge of interest in density estimation in recent years, much of the published research has been concerned with purely technical matters with insufficient emphasis given to the technique's practical value. Furthermore, the subject has been rather inaccessible to the general statistician. The account presented in this book places emphasis on topics of methodological importance, in the hope that this will facilitate broader practical application of density estimation and also encourage research into relevant theoretical work. The book also provides an introduction to the subject for those with general interests in statistics. The important role of density estimation as a graphical technique is reflected by the inclusion of more than 50 graphs and figures throughout the text. Several contexts in which density estimation can be used are discussed, including the exploration and presentation of data, nonparametric discriminant analysis, cluster analysis, simulation and the bootstrap, bump hunting, projection pursuit, and the estimation of hazard rates and other quantities that depend on the density. This book includes general survey of methods available for density estimation. The Kernel method, both for univariate and multivariate data, is discussed in detail, with particular emphasis on ways of deciding how much to smooth and on computation aspects. Attention is also given to adaptive methods, which smooth to a greater degree in the tails of the distribution, and to methods based on the idea of penalized likelihood.

This book deals with parametric and nonparametric density estimation from the maximum (penalized) likelihood point of view, including estimation under constraints. The focal points are existence and uniqueness of the estimators, almost sure convergence rates for the L1 error, and data-driven smoothing parameter selection methods, including their practical performance. The reader will gain insight into technical tools from probability theory and applied mathematics.

This book considers large and challenging multistage decision problems, which can be solved in principle by dynamic programming (DP), but their exact solution is computationally intractable. We discuss solution methods that rely on approximations to produce suboptimal policies with adequate performance. These methods are collectively known by several essentially equivalent names: reinforcement learning, approximate dynamic programming, neuro-dynamic programming. They have been at the forefront of research for the last 25 years, and they underlie, among others, the recent impressive successes of self-learning in the context of games such as chess and Go. Our subject has benefited greatly from the interplay of ideas from optimal control and from artificial intelligence, as it relates to reinforcement learning and simulation-based neural network methods. One of the aims of the book is to explore the common boundary between these two fields and to form a bridge that is accessible by workers with background in either field. Another aim is to organize coherently the broad mosaic of methods that have proved successful in practice while having a solid theoretical and/or logical foundation. This may help researchers and practitioners to find their way through the maze of competing ideas that constitute the current state of the art. This book relates to several of our other books: Neuro-Dynamic Programming (Athena Scientific, 1996), Dynamic Programming and Optimal Control (4th edition, Athena Scientific, 2017), Abstract Dynamic Programming (2nd edition, Athena Scientific, 2018), and Nonlinear Programming (Athena Scientific, 2016). However, the mathematical style of this book is somewhat different. While we provide a rigorous, albeit short, mathematical account of the theory of finite and infinite horizon dynamic programming, and some fundamental approximation methods, we rely more on intuitive explanations and less on proof-based insights. Moreover, our mathematical requirements are quite modest: calculus, a minimal use of matrix-vector algebra, and elementary probability (mathematically complicated arguments involving laws of large numbers and stochastic convergence are bypassed in favor of intuitive explanations). The book illustrates the methodology with many examples and illustrations, and uses a gradual expository approach, which proceeds along four directions: (a) From exact DP to approximate DP: We first discuss exact DP algorithms, explain why they may be difficult to implement, and then use them as the basis for approximations. (b) From finite horizon to infinite horizon problems: We first discuss finite horizon exact and approximate DP methodologies, which are intuitive and mathematically simple, and then progress to infinite horizon problems. (c) From deterministic to stochastic models: We often discuss separately deterministic and stochastic problems, since deterministic problems are simpler and offer special advantages for some of our methods. (d) From model-based to model-free implementations: We first discuss model-based implementations, and then we identify schemes that can be appropriately modified to work with a simulator. The book is related and supplemented by the companion research monograph Rollout, Policy Iteration, and Distributed Reinforcement Learning (Athena Scientific, 2020), which focuses more closely on several topics related to rollout, approximate policy iteration, multiagent problems, discrete and Bayesian optimization, and distributed computation, which are either discussed in less detail or not covered at all in the present book. The author's website contains class notes, and a series of videolectures and slides from a 2021 course at ASU, which address a selection of topics from both books.

Nonparametric Functional Estimation is a compendium of papers, written by experts, in the area of nonparametric functional estimation. This book attempts to be exhaustive in nature and is written both for specialists in the area as well as for students of statistics taking courses at the postgraduate level. The main emphasis throughout the book is on the discussion of several

methods of estimation and on the study of their large sample properties. Chapters are devoted to topics on estimation of density and related functions, the application of density estimation to classification problems, and the different facets of estimation of distribution functions. Statisticians and students of statistics and engineering will find the text very useful.
An applied treatment of the key methods and state-of-the-art tools for visualizing and understanding statistical data Smoothing of Multivariate Data provides an illustrative and hands-on approach to the multivariate aspects of density estimation, emphasizing the use of visualization tools. Rather than outlining the theoretical concepts of classification and regression, this book focuses on the procedures for estimating a multivariate distribution via smoothing. The author first provides an introduction to various visualization tools that can be used to construct representations of multivariate functions, sets, data, and scales of multivariate density estimates. Next, readers are presented with an extensive review of the basic mathematical tools that are needed to asymptotically analyze the behavior of multivariate density estimators, with coverage of density classes, lower bounds, empirical processes, and manipulation of density estimates. The book concludes with an extensive toolbox of multivariate density estimators, including anisotropic kernel estimators, minimization estimators, multivariate adaptive histograms, and wavelet estimators. A completely interactive experience is encouraged, as all examples and figurescan be easily replicated using the R software package, and every chapter concludes with numerous exercises that allow readers to test their understanding of the presented techniques. The R software is freely available on the book's related Web site along with "Code" sections for each chapter that provide short instructions for working in the R environment. Combining mathematical analysis with practical implementations, Smoothing of Multivariate Data is an excellent book for courses in multivariate analysis, data analysis, and nonparametric statistics at the upper-undergraduate and graduatelevels. It also serves as a valuable reference for practitioners and researchers in the fields of statistics, computer science, economics, and engineering.
Written to convey an intuitive feel for both theory and practice, its main objective is to illustrate what a powerful tool density estimation can be when used not only with univariate and bivariate data but also in the higher dimensions of trivariate and quadrivariate information. Major concepts are presented in the context of a histogram in order to simplify the treatment of advanced estimators. Features 12 four-color plates, numerous graphic illustrations as well as a multitude of problems and solutions.
The statistical and mathematical principles of smoothing with a focus on applicable techniques are presented in this book. It naturally splits into two parts: The first part is intended for undergraduate students majoring in mathematics, statistics, econometrics or biometrics whereas the second part is intended to be used by master and PhD students or researchers. The material is easy to accomplish since the e-book character of the text gives a maximum of flexibility in learning (and teaching) intensity.
Kernel smoothing refers to a general methodology for recovery of underlying structure in data sets. The basic principle is that local averaging or smoothing is performed with respect to a kernel function. This book provides uninitiated readers with a feeling for the principles, applications, and analysis of kernel smoothers. This is facilitated by the authors' focus on the simplest settings, namely density estimation and nonparametric regression. They pay particular attention to the problem of choosing the smoothing parameter of a kernel smoother, and also treat the multivariate case in detail. Kernal Smoothing is self-contained and assumes only a basic knowledge of statistics, calculus, and matrix algebra. It is an invaluable introduction to the main ideas of kernel estimation for students and researchers from other discipline and provides a comprehensive reference for those familiar with the topic.
The volume presents innovations in data analysis and classification and gives an overview of the state of the art in these scientific fields and applications. Areas that receive considerable

attention in the book are discrimination and clustering, data analysis and statistics, as well as applications in marketing, finance, and medicine. The reader will find material on recent technical and methodological developments and a large number of applications demonstrating the usefulness of the newly developed techniques.

Focussing on applications, this book covers a very broad range, including simple and complex univariate and multivariate density estimation, nonparametric regression estimation, categorical data smoothing, and applications of smoothing to other areas of statistics. It will thus be of particular interest to data analysts, as arguments generally proceed from actual data rather than statistical theory, while the "Background Material" sections will interest statisticians studying the field. Over 750 references allow researchers to find the original sources for more details, and the "Computational Issues" sections provide sources for statistical software that use the methods discussed. Each chapter includes exercises with a heavily computational focus based upon the data sets used in the book, making it equally suitable as a textbook for a course in smoothing.

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.

Density Estimation for Statistics and Data AnalysisRoutledge

First of all, we want to congratulate two new research communities from M- ico and Brazil that have recently joined the Iberoamerican community and the International Association for Pattern Recognition. We believe that the series of congresses that started as the “Taller Iberoamericano de Reconocimiento de Patrones (TIARP)”, and later became the “Iberoamerican Congress on Pattern Recognition (CIARP)”, has contributed to these groupconsolidatione?orts. We hope that in the near future all the Iberoamerican countries will have their own groups and associations to promote our areas of interest; and that these congresses will serve as the forum for scienti?c research exchange, sharing of - pertise and new knowledge, and establishing contacts that improve cooperation between research groups in pattern recognition and related areas. CIARP 2004

(9th Iberoamerican Congress on Pattern Recognition) was the ninthinaseriesofpioneeringcongressesonpatternrecognitionintheIberoam- ican community. As in the previous year, CIARP 2004 also included worldwide participation. It took place in Puebla, Mexico. The aim of the congress was to promote and disseminate ongoing research and mathematical methods for pattern recognition, image analysis, and applications in such diverse areas as computer vision, robotics, industry, health, entertainment, space exploration, telecommunications, data mining, document analysis,and natural languagep- cessing and recognition, to name a few.

Modern analysis of HEP data needs advanced statistical tools to separate signal from background. This is the first book which focuses on machine learning techniques. It will be of interest to almost every high energy physicist, and, due to its coverage, suitable for students.

Over the last forty years there has been a growing interest to extend probability theory and statistics and to allow for more flexible modelling of imprecision, uncertainty, vagueness and ignorance. The fact that in many real-life situations data uncertainty is not only present in the form of randomness (stochastic uncertainty) but also in the form of imprecision/fuzziness is but one point underlining the need for a widening of statistical tools. Most such extensions originate in a "softening" of classical methods, allowing, in particular, to work with imprecise or vague data, considering imprecise or generalized probabilities and fuzzy events, etc. About ten years ago the idea of establishing a recurrent forum for discussing new trends in the before-mentioned context was born and resulted in the first International Conference on Soft Methods in Probability and Statistics (SMPS) that was held in Warsaw in 2002. In the following years the conference took place in Oviedo (2004), in Bristol (2006) and in Toulouse (2008). In the current edition the conference returns to Oviedo. This edited volume is a collection of papers presented at the SMPS 2010 conference held in Mieres and Oviedo. It gives a comprehensive overview of current research into the fusion of soft methods with probability and statistics.

Clarifies modern data analysis through nonparametric density estimation for a complete working knowledge of the theory and methods Featuring a thoroughly revised presentation, Multivariate Density Estimation: Theory, Practice, and Visualization, Second Edition maintains an intuitive approach to the underlying methodology and supporting theory of density estimation. Including new material and updated research in each chapter, the Second Edition presents additional clarification of theoretical opportunities, new algorithms, and up-to-date coverage of the unique challenges presented in the field of data analysis. The new edition focuses on the various density estimation techniques and methods that can be used in the field of big data. Defining optimal nonparametric estimators, the Second Edition demonstrates the density estimation tools to use when dealing with various multivariate structures in univariate, bivariate, trivariate, and quadrivariate data analysis. Continuing to illustrate the major concepts in the context of the classical histogram, Multivariate Density Estimation: Theory, Practice, and Visualization, Second Edition also features: Over 150 updated figures to clarify theoretical results and to show analyses of real data sets An updated presentation of graphic visualization using computer software such as R A clear discussion of selections of important research during the past decade, including mixture estimation, robust parametric modeling algorithms, and clustering More than 130 problems to help readers reinforce the main concepts and ideas presented Boxed theorems and results allowing easy identification of crucial ideas Figures in color in the digital versions of the book A website with related data sets Multivariate Density Estimation: Theory, Practice, and Visualization, Second Edition is an ideal reference for theoretical and applied statisticians, practicing engineers, as well as readers interested in the theoretical aspects of nonparametric estimation and the application of these methods to multivariate data. The Second Edition is also useful as a textbook for introductory courses in kernel statistics, smoothing, advanced computational statistics, and general forms of statistical distributions.

This book gives a rigorous, systematic treatment of density estimates, their construction, use and analysis with full proofs. It develops L1 theory, rather than the classical L2, showing how L1 exposes fundamental properties of density estimates masked by L2.

Density estimation has evolved enormously since the days of bar plots and histograms, but researchers and users are still struggling with the problem of the selection of the bin widths. This book is the first to explore a new paradigm for the data-based or automatic selection of the free parameters of density estimates in general so that the expected error is within a given constant multiple of the best possible error. The paradigm can be used in nearly all density estimates and for most model selection problems, both parametric and nonparametric.

Measuring the degree of association between random variables is a task inherent in many practical applications such as risk management and financial modeling. Well-known measures like Spearman's rho and Kendall's tau can be expressed in terms of the underlying copula only, hence, being independent of the underlying univariate marginal distributions. Opposed to these classical measures of association, mutual information, which is derived from information theory, constitutes a fundamentally different approach of measuring association. Although this measure is likewise independent of the univariate margins, it is not a functional of the copula but of the corresponding copula density. Besides the theoretical properties of mutual information as a measure of multivariate association, possibilities to estimate the copula density based on observations of continuous distributions are investigated. To cope with the effect of boundary bias, new estimators are introduced and existing functionals are generalized to the multivariate case. The performance of these estimators is evaluated in comparison to common kernel density estimation schemes. To facilitate variance estimation by means of resampling methods like bootstrapping, an algorithm is introduced, which significantly reduces computation time in comparison with pre-implemented algorithms. In practical applications, complete continuous data is oftentimes not available to the analyst. Instead, categorial data derived from the underlying continuous distribution may be given. Hence, estimation of the copula and its density based on contingency tables is investigated. The newly developed estimators are employed to derive estimates of Spearman's rho and Kendall's tau and their performance is compared.

Solve real-world data problems with R and machine learning Key Features Third edition of the bestselling, widely acclaimed R machine learning book, updated and improved for R 3.6 and beyond Harness the power of R to build flexible, effective, and transparent machine learning models Learn quickly with a clear, hands-on guide by experienced machine learning teacher and practitioner, Brett Lantz Book Description Machine learning, at its core, is concerned with transforming data into actionable knowledge. R offers a powerful set of machine learning methods to quickly and easily gain insight from your data. Machine Learning with R, Third Edition provides a hands-on, readable guide to applying machine learning to real-world problems. Whether you are an experienced R user or new to the language, Brett Lantz teaches you everything you need to uncover key insights, make new predictions, and visualize your findings. This new 3rd edition updates the classic R data science book to R 3.6 with newer and better libraries, advice on ethical and bias issues in machine learning, and an introduction to deep learning. Find powerful new insights in your data; discover machine learning with R. What you will learn Discover the origins of machine learning and how exactly a computer learns by example Prepare your data for machine learning work with the R programming language Classify important outcomes using nearest neighbor and Bayesian methods Predict future events using decision trees, rules, and support vector machines Forecast numeric data and estimate financial values using regression methods Model complex processes with artificial neural networks — the basis of deep learning Avoid bias in machine learning models Evaluate your models and improve their performance Connect R to SQL databases and emerging big data technologies such as Spark, H2O, and TensorFlow Who this book is for Data scientists, students, and other practitioners who want a clear, accessible guide to machine learning with R.

This text provides the reader with a single book where they can find accounts of a number of up-to-date issues in nonparametric inference. The book is aimed at Masters or PhD level students in statistics, computer science, and engineering. It is also suitable for researchers who want to get up to speed quickly on modern nonparametric methods. It covers a wide range of topics including the bootstrap, the nonparametric delta method, nonparametric regression, density estimation, orthogonal function methods, minimax estimation, nonparametric confidence sets, and wavelets. The book's dual approach includes a mixture of methodology and theory.

This book presents the details of the BONUS algorithm and its real world applications in areas like sensor placement in large scale drinking water networks, sensor placement in advanced power systems, water management in power systems, and capacity expansion of energy systems. A generalized method for stochastic nonlinear programming based on a sampling based approach for uncertainty analysis and statistical reweighting to obtain probability information is demonstrated in this book. Stochastic optimization problems are difficult to solve since they involve dealing with optimization and uncertainty loops. There are two fundamental approaches used to solve such problems. The first being the decomposition techniques and the second method identifies problem specific structures and transforms the problem into a deterministic nonlinear programming problem. These techniques have significant limitations on either the objective function type or the underlying distributions for the uncertain variables. Moreover, these methods assume that there are a small number of scenarios to be evaluated for calculation of the probabilistic objective function and constraints. This book begins to tackle these issues by describing a generalized method for stochastic nonlinear programming problems. This title is best suited for practitioners, researchers and students in engineering, operations research, and management science who desire a complete understanding of the BONUS algorithm and its applications to the real world.

For many researchers, Python is a first-class tool mainly because of its libraries for storing, manipulating, and gaining insight from data. Several resources exist for individual pieces of this data science stack, but only with the Python Data Science Handbook do you get them all—IPython, NumPy, Pandas, Matplotlib, Scikit-Learn, and other related tools. Working scientists and data crunchers familiar with reading and writing Python code will find this comprehensive desk reference ideal for tackling day-to-day issues: manipulating, transforming, and cleaning data; visualizing different types of data; and using data to build statistical or machine learning models. Quite simply, this is the must-have reference for scientific computing in Python. With this handbook, you'll learn how to use: IPython and Jupyter: provide computational environments for data scientists using Python NumPy: includes the ndarray for efficient storage and manipulation of dense data arrays in Python Pandas: features the DataFrame for efficient storage and manipulation of labeled/columnar data in Python Matplotlib: includes capabilities for a flexible range of data visualizations in Python Scikit-Learn: for efficient and clean Python implementations of the most important and established machine learning algorithms

This authoritative new volume treats a wide class of distributions that constitute plausible alternatives to normality -- such as short- and long-tailed symmetric distributions and moderately skewed distributions -- all having finite mean and variance. Robust Inference illustrates the appropriateness of various robust methods for solving both one-sample and multisample statistical inference problems ... develops Laguerre series expansions for Student's t and variance-ratio F statistic distributions ... analyzes normal and nonnormal distribution efficiencies ... works out modified maximum likelihood (MML) estimators based on type II censored samples for log-normal, logistic, exponential, and Rayleigh distributions ... uses MML estimators in constructing robust hypothesis-testing procedures ... considers the specialized topics of regression, analysis of variance, classification, and sample survey ... discusses goodness-of-fit tests ... describes Q-Q plots in a special appendix ... and much more. An outstanding, time-saving reference for theoreticians and practitioners of statistics, Robust Inference is also an excellent auxiliary text for an undergraduate- or graduate-level course on robustness. Book jacket.

Until now, students and researchers in nonparametric and semiparametric statistics and econometrics have had to turn to the latest journal articles to keep pace with these emerging methods of economic analysis. Nonparametric Econometrics fills a major gap by gathering together the most up-to-date theory and techniques and presenting them in a remarkably straightforward and accessible format. The empirical tests, data, and exercises included in this textbook help make it the ideal introduction for graduate students and an indispensable resource for researchers. Nonparametric and semiparametric methods have attracted a great deal of attention from statisticians in recent decades. While the majority of existing books on the subject operate from the presumption that the underlying data is strictly continuous in nature, more often than not social scientists deal with categorical data--nominal and ordinal--in applied settings. The conventional nonparametric approach to dealing with the presence of discrete variables is acknowledged to be unsatisfactory. This book is tailored to the needs of applied econometricians and social scientists. Qi Li and Jeffrey Racine emphasize nonparametric techniques suited to the rich array of data types--continuous, nominal, and ordinal--within one coherent framework. They also emphasize the properties of nonparametric estimators in the presence of potentially irrelevant variables. Nonparametric Econometrics covers all the material necessary to understand and apply nonparametric methods for real-world problems.

Probability is the bedrock of machine learning. You cannot develop a deep understanding and application of machine learning without it. Cut through the equations, Greek letters, and confusion, and discover the topics in probability that you need to know. Using clear explanations, standard Python libraries, and step-by-step tutorial lessons, you will discover the importance of probability to machine learning, Bayesian probability, entropy, density estimation, maximum likelihood, and much more.

The First Asian Conference on Machine Learning (ACML 2009) was held at Nanjing, China during November 2–4, 2009.This was the ?rst edition of a series of annual conferences which aim to provide a leading international forum for researchers in machine learning and related ?elds to share their new ideas and research ?ndings. This year we received 113 submissions from 18 countries and regions in Asia, Australasia, Europe and North America. The submissions went through a r- orous double-blind reviewing process. Most submissions received four reviews, a few submissions received ?ve reviews, while only several submissions received three reviews. Each submission was handled by an Area Chair who coordinated discussions among reviewers and made recommendation on the submission. The Program Committee Chairs examined the reviews and meta-reviews to further guarantee the reliability and integrity of the reviewing process. Twenty-nine - pers were selected after this process. To ensure that important revisions required by reviewers were incorporated into the ?nal accepted papers, and to allow submissions which would have - tential after a careful revision, this year we launched a “revision double-check” process. In short, the above-mentioned 29 papers were conditionally accepted, and the authors were requested to incorporate the “important-and-must”re- sionssummarizedbyareachairsbasedonreviewers’comments.Therevised?nal version and the revision list of each conditionally accepted paper was examined by the Area Chair and Program Committee Chairs. Papers that failed to pass the

examination were ?nally rejected.
A coherent introductory text from a groundbreaking researcher, focusing on clarity and motivation to build intuition and understanding.
Kernel smoothing has greatly evolved since its inception to become an essential methodology in the data science tool kit for the 21st century. Its widespread adoption is due to its fundamental role for multivariate exploratory data analysis, as well as the crucial role it plays in composite solutions to complex data challenges. Multivariate Kernel Smoothing and Its Applications offers a comprehensive overview of both aspects. It begins with a thorough exposition of the approaches to achieve the two basic goals of estimating probability density functions and their derivatives. The focus then turns to the applications of these approaches to more complex data analysis goals, many with a geometric/topological flavour, such as level set estimation, clustering (unsupervised learning), principal curves, and feature significance. Other topics, while not direct applications of density (derivative) estimation but sharing many commonalities with the previous settings, include classification (supervised learning), nearest neighbour estimation, and deconvolution for data observed with error. For a data scientist, each chapter contains illustrative Open data examples that are analysed by the most appropriate kernel smoothing method. The emphasis is always placed on an intuitive understanding of the data provided by the accompanying statistical visualisations. For a reader wishing to investigate further the details of their underlying statistical reasoning, a graduated exposition to a unified theoretical framework is provided. The algorithms for efficient software implementation are also discussed. José E. Chacón is an associate professor at the Department of Mathematics of the Universidad de Extremadura in Spain. Tarn Duong is a Senior Data Scientist for a start-up which provides short distance carpooling services in France. Both authors have made important contributions to kernel smoothing research over the last couple of decades.