Applied Predictive Modeling Max Kuhn Ajban

Bayesian Modeling Using WinBUGSMachine Learning with RMachine Learning EssentialsMachine Learning with RApplied Analytics through Case Studies Using SAS and RApplied Machine LearningModeling Discrete Time-to-Event DataApplied Predictive ModelingDynamic Documents with R and knitrFeature Engineering and SelectionMastering Predictive Analytics with RLearning Predictive Analytics with RPredictive AnalyticsPractical R for Mass Communication and JournalismThe Art of Feature EngineeringCalculated RisksMachine Learning for HackersNonclinical Statistics for Pharmaceutical and Biotechnology IndustriesA Modern Approach to Regression with RRegression Modeling StrategiesPredictive Analytics For DummiesBayesian NetworksStatistical Learning from a Regression PerspectiveApplied Predictive AnalyticsHands-On Machine Learning with RApplied Predictive ModelingInterpretable Machine LearningHands-On Predictive Analytics with PythonLearning from Imbalanced Data SetsPredictive Analytics, Data Mining and Big DataSentence logicAn Introduction to Stochastic DynamicsData Mining and Predictive AnalyticsStatistical Inference via Data Science: A ModernDive into R and the TidyverseMachine Learning with RModeling to Inform Infectious Disease ControlLearning Predictive Analytics with PythonStatistical Modeling and ComputationMarketing Data ScienceFeature Engineering for Machine Learning

Bayesian Modeling Using WinBUGS

Applied Predictive Modeling covers the overall predictive modeling process, beginning with the crucial steps of data preprocessing, data splitting and foundations of model tuning. The text then provides intuitive explanations of numerous common and modern regression and classification techniques, always with an emphasis on illustrating and solving real data problems. The text illustrates all parts of the modeling process through many hands-on, real-life examples, and every chapter contains extensive R code for each step of the process. This multi-purpose text can be used as an introduction to predictive models and the overall modeling process, a practitioner's reference handbook, or as a text for advanced undergraduate or graduate level predictive modeling courses. To that end, each chapter contains problem sets to help solidify the covered concepts and uses data available in the book's R package. This text is intended for a broad audience as both an introduction to predictive models as well as a guide to applying them. Non-mathematical readers will appreciate the intuitive explanations of the techniques while an emphasis on problem-solving with real data across a wide variety of applications will aid practitioners who wish to extend their expertise. Readers should have knowledge of basic statistical ideas, such as correlation and linear regression analysis. While the text is biased against complex equations, a mathematical background is needed for advanced topics.

Machine Learning with R

Do you want to use R to tell stories? This book was written for you—whether you already know some R or have never coded before. Most R texts focus only on programming or statistical theory. Practical R for Mass Communication and Journalism gives you ideas, tools, and techniques for incorporating data and visualizations into your narratives. You'll see step by step how to: Analyze airport flight delays, restaurant inspections, and election results Map bank locations, median incomes, and new voting districts Compare campaign contributions to final election results Extract data from PDFs Whip messy data into shape for analysis Scrape data from a website Create graphics ranging from simple, static charts to interactive visualizations for the Web If you work or plan to work in a newsroom, government office, non-profit policy organization, or PR office, Practical R for Mass Communication and Journalism will help you use R in your world. This book has a companion website with code, links to additional resources, and searchable tables by function and task. Sharon Machlis is the author of Computerworld's Beginner's Guide to R, host of InfoWorld's Do More With R video screencast series, admin for the R for Journalists Google Group, and is well known among Twitter users who follow the #rstats hashtag. She is Director of Editorial Data and Analytics at IDG Communications (parent company of Computerworld, InfoWorld, PC World and Macworld, among others) and a frequent speaker at data journalism and R conferences.

Machine Learning Essentials

Written as a tutorial to explore and understand the power of R for machine learning. This practical guide that covers all of the need to know topics in a very systematic way. For each machine learning approach, each step in the process is detailed, from preparing the data for analysis to evaluating the results. These steps will build the knowledge you need to apply them to your own data science tasks.Intended for those who want to learn how to use R's machine learning capabilities and gain insight from your data. Perhaps you already know a bit about machine learning, but have never used R; or perhaps you know a little R but are new to machine learning. In either case, this book will get you up and running quickly. It would be helpful to have a bit of familiarity with basic programming concepts, but no prior experience is required.

Machine Learning with R

Combine business sense, statistics, and computers in a new and intuitive way, thanks to Big Data Predictive analytics is a branch of data mining that helps predict probabilities and trends. Predictive Analytics For Dummies explores the power of predictive analytics and how you can use it to make valuable predictions for your business, or in fields such as advertising, fraud detection, politics, and others. This

practical book does not bog you down with loads of mathematical or scientific theory, but instead helps you quickly see how to use the right algorithms and tools to collect and analyze data and apply it to make predictions. Topics include using structured and unstructured data, building models, creating a predictive analysis roadmap, setting realistic goals, budgeting, and much more. Shows readers how to use Big Data and data mining to discover patterns and make predictions for tech-savvy businesses Helps readers see how to shepherd predictive analytics projects through their companies Explains just enough of the science and math, but also focuses on practical issues such as protecting project budgets, making good presentations, and more Covers nuts-and-bolts topics including predictive analytics basics, using structured and unstructured data, data mining, and algorithms and techniques for analyzing data Also covers clustering, association, and statistical models; creating a predictive analytics roadmap; and applying predictions to the web, marketing, finance, health care, and elsewhere Propose, produce, and protect predictive analytics projects through your company with Predictive Analytics For Dummies.

Applied Analytics through Case Studies Using SAS and R

At the beginning of the twentieth century, H. G. Wells predicted that statistical thinking would be as necessary for citizenship in a technological world as the ability to read and write. But in the twenty-first century, we are often overwhelmed

by a baffling array of percentages and probabilities as we try to navigate in a world dominated by statistics. Cognitive scientist Gerd Gigerenzer says that because we haven't learned statistical thinking, we don't understand risk and uncertainty. In order to assess risk -- everything from the risk of an automobile accident to the certainty or uncertainty of some common medical screening tests -- we need a basic understanding of statistics. Astonishingly, doctors and lawyers don't understand risk any better than anyone else. Gigerenzer reports a study in which doctors were told the results of breast cancer screenings and then were asked to explain the risks of contracting breast cancer to a woman who received a positive result from a screening. The actual risk was small because the test gives many false positives. But nearly every physician in the study overstated the risk. Yet many people will have to make important health decisions based on such information and the interpretation of that information by their doctors. Gigerenzer explains that a major obstacle to our understanding of numbers is that we live with an illusion of certainty. Many of us believe that HIV tests, DNA fingerprinting, and the growing number of genetic tests are absolutely certain. But even DNA evidence can produce spurious matches. We cling to our illusion of certainty because the medical industry, insurance companies, investment advisers, and election campaigns have become purveyors of certainty, marketing it like a commodity. To avoid confusion, says Gigerenzer, we should rely on more understandable representations of risk, such as absolute risks. For example, it is said that a mammography screening reduces the risk of breast cancer by 25 percent. But in

absolute risks, that means that out of every 1,000 women who do not participate in screening, 4 will die; while out of 1,000 women who do, 3 will die. A 25 percent risk reduction sounds much more significant than a benefit that 1 out of 1,000 women will reap. This eye-opening book explains how we can overcome our ignorance of numbers and better understand the risks we may be taking with our money, our health, and our lives.

Applied Machine Learning

Get to grips with key data visualization and predictive analytic skills using R About This Book Acquire predictive analytic skills using various tools of R Make predictions about future events by discovering valuable information from data using R Comprehensible guidelines that focus on predictive model design with real-world data Who This Book Is For If you are a statistician, chief information officer, data scientist, ML engineer, ML practitioner, quantitative analyst, and student of machine learning, this is the book for you. You should have basic knowledge of the use of R. Readers without previous experience of programming in R will also be able to use the tools in the book. What You Will Learn Customize R by installing and loading new packages Explore the structure of data using clustering algorithms Turn unstructured text into ordered data, and acquire knowledge from the data Classify your observations using Naive Bayes, k-NN, and decision trees Reduce the dimensionality of your data using principal component analysis

Discover association rules using Apriori Understand how statistical distributions can help retrieve information from data using correlations, linear regression, and multilevel regression Use PMML to deploy the models generated in R In Detail R is statistical software that is used for data analysis. There are two main types of learning from data: unsupervised learning, where the structure of data is extracted automatically; and supervised learning, where a labeled part of the data is used to learn the relationship or scores in a target attribute. As important information is often hidden in a lot of data, R helps to extract that information with its many standard and cutting-edge statistical functions. This book is packed with easy-to-follow guidelines that explain the workings of the many key data mining tools of R, which are used to discover knowledge from your data. You will learn how to perform key predictive analytics tasks using R, such as train and test predictive models for classification and regression tasks, score new data sets and so on. All chapters will guide you in acquiring the skills in a practical way. Most chapters also include a theoretical introduction that will sharpen your understanding of the subject matter and invite you to go further. The book familiarizes you with the most common data mining tools of R, such as k-means, hierarchical regression, linear regression, association rules, principal component analysis, multilevel modeling, k-NN, Naive Bayes, decision trees, and text mining. It also provides a description of visualization techniques using the basic visualization tools of R as well as lattice for visualizing patterns in data organized in groups. This book is invaluable for anyone fascinated by the data mining opportunities offered by GNU

R and its packages. Style and approach This is a practical book, which analyzes compelling data about life, health, and death with the help of tutorials. It offers you a useful way of interpreting the data that's specific to this book, but that can also be applied to any other data.

Modeling Discrete Time-to-Event Data

This textbook on statistical modeling and statistical inference will assist advanced undergraduate and graduate students. Statistical Modeling and Computation provides a unique introduction to modern Statistics from both classical and Bayesian perspectives. It also offers an integrated treatment of Mathematical Statistics and modern statistical computation, emphasizing statistical modeling, computational techniques, and applications. Each of the three parts will cover topics essential to university courses. Part I covers the fundamentals of probability theory. In Part II, the authors introduce a wide variety of classical models that include, among others, linear regression and ANOVA models. In Part III, the authors address the statistical analysis and computation of various advanced models, such as generalized linear, state-space and Gaussian models. Particular attention is paid to fast Monte Carlo techniques for Bayesian inference on these models. Throughout the book the authors include a large number of illustrative examples and solved problems. The book also features a section with solutions, an appendix that serves as a MATLAB primer, and a mathematical supplement.

Applied Predictive Modeling

This textbook considers statistical learning applications when interest centers on the conditional distribution of a response variable, given a set of predictors, and in the absence of a credible model that can be specified before the data analysis begins. Consistent with modern data analytics, it emphasizes that a proper statistical learning data analysis depends in an integrated fashion on sound data collection, intelligent data management, appropriate statistical procedures, and an accessible interpretation of results. The unifying theme is that supervised learning properly can be seen as a form of regression analysis. Key concepts and procedures are illustrated with a large number of real applications and their associated code in R, with an eye toward practical implications. The growing integration of computer science and statistics is well represented including the occasional, but salient, tensions that result. Throughout, there are links to the big picture. The third edition considers significant advances in recent years, among which are: the development of overarching, conceptual frameworks for statistical learning; the impact of "big data" on statistical learning; the nature and consequences of post-model selection statistical inference; deep learning in various forms; the special challenges to statistical inference posed by statistical learning; the fundamental connections between data collection and data analysis; interdisciplinary ethical and political issues surrounding the application of algorithmic methods in a wide variety of fields, each linked to concerns about

transparency, fairness, and accuracy. This edition features new sections on accuracy, transparency, and fairness, as well as a new chapter on deep learning. Precursors to deep learning get an expanded treatment. The connections between fitting and forecasting are considered in greater depth. Discussion of the estimation targets for algorithmic methods is revised and expanded throughout to reflect the latest research. Resampling procedures are emphasized. The material is written for upper undergraduate and graduate students in the social, psychological and life sciences and for researchers who want to apply statistical learning procedures to scientific and policy problems.

Dynamic Documents with R and knitr

This book focuses on tools and techniques for building regression models using real-world data and assessing their validity. A key theme throughout the book is that it makes sense to base inferences or conclusions only on valid models. Plots are shown to be an important tool for both building regression models and assessing their validity. We shall see that deciding what to plot and how each plot should be interpreted will be a major challenge. In order to overcome this challenge we shall need to understand the mathematical properties of the fitted regression models and associated diagnostic procedures. As such this will be an area of focus throughout the book. In particular, we shall carefully study the properties of resi-als in order to understand when patterns in residual plots provide direct

information about model misspecification and when they do not. The regression output and plots that appear throughout the book have been gen- ated using R. The output from R that appears in this book has been edited in minor ways. On the book web site you will find the R code used in each example in the text.

Feature Engineering and Selection

Gain practical insights into predictive modelling by implementing Predictive Analytics algorithms on public datasets with Python About This Book A step-by-step guide to predictive modeling including lots of tips, tricks, and best practices Get to grips with the basics of Predictive Analytics with Python Learn how to use the popular predictive modeling algorithms such as Linear Regression, Decision Trees, Logistic Regression, and Clustering Who This Book Is For If you wish to learn how to implement Predictive Analytics algorithms using Python libraries, then this is the book for you. If you are familiar with coding in Python (or some other programming/statistical/scripting language) but have never used or read about Predictive Analytics algorithms, this book will also help you. The book will be beneficial to and can be read by any Data Science enthusiasts. Some familiarity with Python will be useful to get the most out of this book, but it is certainly not a prerequisite. What You Will Learn Understand the statistical and mathematical concepts behind Predictive Analytics algorithms and implement Predictive Analytics algorithms using Python libraries Analyze the result parameters arising

from the implementation of Predictive Analytics algorithms Write Python modules/functions from scratch to execute segments or the whole of these algorithms Recognize and mitigate various contingencies and issues related to the implementation of Predictive Analytics algorithms Get to know various methods of importing, cleaning, sub-setting, merging, joining, concatenating, exploring, grouping, and plotting data with pandas and numpy Create dummy datasets and simple mathematical simulations using the Python numpy and pandas libraries Understand the best practices while handling datasets in Python and creating predictive models out of them In Detail Social Media and the Internet of Things have resulted in an avalanche of data. Data is powerful but not in its raw form - It needs to be processed and modeled, and Python is one of the most robust tools out there to do so. It has an array of packages for predictive modeling and a suite of IDEs to choose from. Learning to predict who would win, lose, buy, lie, or die with Python is an indispensable skill set to have in this data age. This book is your guide to getting started with Predictive Analytics using Python. You will see how to process data and make predictive models from it. We balance both statistical and mathematical concepts, and implement them in Python using libraries such as pandas, scikit-learn, and numpy. You'll start by getting an understanding of the basics of predictive modeling, then you will see how to cleanse your data of impurities and get it ready it for predictive modeling. You will also learn more about the best predictive modeling algorithms such as Linear Regression, Decision Trees, and Logistic Regression. Finally, you will see the best practices in predictive

modeling, as well as the different applications of predictive modeling in the modern world. Style and approach All the concepts in this book been explained and illustrated using a dataset, and in a step-by-step manner. The Python code snippet to implement a method or concept is followed by the output, such as charts, dataset heads, pictures, and so on. The statistical concepts are explained in detail wherever required.

Mastering Predictive Analytics with R

A practical guide for data scientists who want to improve the performance of any machine learning solution with feature engineering.

Learning Predictive Analytics with R

Understand the Foundations of Bayesian Networks—Core Properties and Definitions Explained Bayesian Networks: With Examples in R introduces Bayesian networks using a hands-on approach. Simple yet meaningful examples in R illustrate each step of the modeling process. The examples start from the simplest notions and gradually increase in complexity. The authors also distinguish the probabilistic models from their estimation with data sets. The first three chapters explain the whole process of Bayesian network modeling, from structure learning to parameter

learning to inference. These chapters cover discrete Bayesian, Gaussian Bayesian, and hybrid networks, including arbitrary random variables. The book then gives a concise but rigorous treatment of the fundamentals of Bayesian networks and offers an introduction to causal Bayesian networks. It also presents an overview of R and other software packages appropriate for Bayesian networks. The final chapter evaluates two real-world examples: a landmark causal protein signaling network paper and graphical modeling approaches for predicting the composition of different body parts. Suitable for graduate students and non-statisticians, this text provides an introductory overview of Bayesian networks. It gives readers a clear, practical understanding of the general approach and steps involved.

Predictive Analytics

This book serves as a reference text for regulatory, industry and academic statisticians and also a handy manual for entry level Statisticians. Additionally it aims to stimulate academic interest in the field of Nonclinical Statistics and promote this as an important discipline in its own right. This text brings together for the first time in a single volume a comprehensive survey of methods important to the nonclinical science areas within the pharmaceutical and biotechnology industries. Specifically the Discovery and Translational sciences, the Safety/Toxiology sciences, and the Chemistry, Manufacturing and Controls sciences. Drug discovery and development is a long and costly process. Most

decisions in the drug development process are made with incomplete information. The data is rife with uncertainties and hence risky by nature. This is therefore the purview of Statistics. As such, this book aims to introduce readers to important statistical thinking and its application in these nonclinical areas. The chapters provide as appropriate, a scientific background to the topic, relevant regulatory guidance, current statistical practice, and further research directions.

Practical R for Mass Communication and Journalism

Step-by-step guide to build high performing predictive applications Key Features Use the Python data analytics ecosystem to implement end-to-end predictive analytics projects Explore advanced predictive modeling algorithms with an emphasis on theory with intuitive explanations Learn to deploy a predictive model's results as an interactive application Book Description Predictive analytics is an applied field that employs a variety of quantitative methods using data to make predictions. It involves much more than just throwing data onto a computer to build a model. This book provides practical coverage to help you understand the most important concepts of predictive analytics. Using practical, step-by-step examples, we build predictive analytics solutions while using cutting-edge Python tools and packages. The book's step-by-step approach starts by defining the problem and moves on to identifying relevant data. We will also be performing data preparation, exploring and visualizing relationships, building models, tuning,

evaluating, and deploying model. Each stage has relevant practical examples and efficient Python code. You will work with models such as KNN, Random Forests, and neural networks using the most important libraries in Python's data science stack: NumPy, Pandas, Matplotlib, Seaborn, Keras, Dash, and so on. In addition to hands-on code examples, you will find intuitive explanations of the inner workings of the main techniques and algorithms used in predictive analytics. By the end of this book, you will be all set to build high-performance predictive analytics solutions using Python programming. What you will learn Get to grips with the main concepts and principles of predictive analytics Learn about the stages involved in producing complete predictive analytics solutions Understand how to define a problem, propose a solution, and prepare a dataset Use visualizations to explore relationships and gain insights into the dataset Learn to build regression and classification models using scikit-learn Use Keras to build powerful neural network models that produce accurate predictions Learn to serve a model's predictions as a web application Who this book is for This book is for data analysts, data scientists, data engineers, and Python developers who want to learn about predictive modeling and would like to implement predictive analytics solutions using Python's data stack. People from other backgrounds who would like to enter this exciting field will greatly benefit from reading this book. All you need is to be proficient in Python programming and have a basic understanding of statistics and college-level algebra.

The Art of Feature Engineering

Many texts are excellent sources of knowledge about individual statistical tools, but the art of data analysis is about choosing and using multiple tools. Instead of presenting isolated techniques, this text emphasizes problem solving strategies that address the many issues arising when developing multivariable models using real data and not standard textbook examples. It includes imputation methods for dealing with missing data effectively, methods for dealing with nonlinear relationships and for making the estimation of transformations a formal part of the modeling process, methods for dealing with "too many variables to analyze and not enough observations," and powerful model validation techniques based on the bootstrap. This text realistically deals with model uncertainty and its effects on inference to achieve "safe data mining".

Calculated Risks

Learn the art and science of predictive analytics — techniques that get results Predictive analytics is what translates big data into meaningful, usable business information. Written by a leading expert in the field, this guide examines the science of the underlying algorithms as well as the principles and best practices that govern the art of predictive analytics. It clearly explains the theory behind

predictive analytics, teaches the methods, principles, and techniques for conducting predictive analytics projects, and offers tips and tricks that are essential for successful predictive modeling. Hands-on examples and case studies are included. The ability to successfully apply predictive analytics enables businesses to effectively interpret big data; essential for competition today This guide teaches not only the principles of predictive analytics, but also how to apply them to achieve real, pragmatic solutions Explains methods, principles, and techniques for conducting predictive analytics projects from start to finish Illustrates each technique with hands-on examples and includes as series of in-depth case studies that apply predictive analytics to common business scenarios A companion website provides all the data sets used to generate the examples as well as a free trial version of software Applied Predictive Analytics arms data and business analysts and business managers with the tools they need to interpret and capitalize on big data.

Machine Learning for Hackers

Hands-on Machine Learning with R provides a practical and applied approach to learning and developing intuition into today's most popular machine learning methods. This book serves as a practitioner's guide to the machine learning process and is meant to help the reader learn to apply the machine learning stack within R, which includes using various R packages such as glmnet, h2o, ranger,

xgboost, keras, and others to effectively model and gain insight from their data. The book favors a hands-on approach, providing an intuitive understanding of machine learning concepts through concrete examples and just a little bit of theory. Throughout this book, the reader will be exposed to the entire machine learning process including feature engineering, resampling, hyperparameter tuning, model evaluation, and interpretation. The reader will be exposed to powerful algorithms such as regularized regression, random forests, gradient boosting machines, deep learning, generalized low rank models, and more! By favoring a hands-on approach and using real word data, the reader will gain an intuitive understanding of the architectures and engines that drive these algorithms and packages, understand when and how to tune the various hyperparameters, and be able to interpret model results. By the end of this book, the reader should have a firm grasp of R's machine learning stack and be able to implement a systematic approach for producing high quality modeling results. Features: · Offers a practical and applied introduction to the most popular machine learning methods. · Topics covered include feature engineering, resampling, deep learning and more. · Uses a hands-on approach and real world data.

Nonclinical Statistics for Pharmaceutical and Biotechnology Industries

Discovering knowledge from big multivariate data, recorded every days, requires specialized machine learning techniques. This book presents an easy to use practical guide in R to compute the most popular machine learning methods for exploring real word data sets, as well as, for building predictive models. The main parts of the book include: A) Unsupervised learning methods, to explore and discover knowledge from a large multivariate data set using clustering and principal component methods. You will learn hierarchical clustering, k-means, principal component analysis and correspondence analysis methods. B) Regression analysis, to predict a quantitative outcome value using linear regression and non-linear regression strategies. C) Classification techniques, to predict a qualitative outcome value using logistic regression, discriminant analysis, naive bayes classifier and support vector machines. D) Advanced machine learning methods, to build robust regression and classification models using k-nearest neighbors methods, decision tree models, ensemble methods (bagging, random forest and boosting). E) Model selection methods, to select automatically the best combination of predictor variables for building an optimal predictive model. These include, best subsets selection methods, stepwise regression and penalized regression (ridge, lasso and elastic net regression models). We also present principal component-based regression methods, which are useful when the data contain multiple correlated predictor variables. F) Model validation and evaluation techniques for measuring the performance of a predictive model. G) Model diagnostics for detecting and fixing a potential problems in a predictive model. The

book presents the basic principles of these tasks and provide many examples in R. This book offers solid guidance in data mining for students and researchers. Key features: - Covers machine learning algorithm and implementation - Key mathematical concepts are presented - Short, self-contained chapters with practical examples.

A Modern Approach to Regression with R

Feature engineering is a crucial step in the machine-learning pipeline, yet this topic is rarely examined on its own. With this practical book, you'll learn techniques for extracting and transforming features—the numeric representations of raw data—into formats for machine-learning models. Each chapter guides you through a single data problem, such as how to represent text or image data. Together, these examples illustrate the main principles of feature engineering. Rather than simply teach these principles, authors Alice Zheng and Amanda Casari focus on practical application with exercises throughout the book. The closing chapter brings everything together by tackling a real-world, structured dataset with several feature-engineering techniques. Python packages including numpy, Pandas, Scikit-learn, and Matplotlib are used in code examples. You'll examine: Feature engineering for numeric data: filtering, binning, scaling, log transforms, and power transforms Natural text techniques: bag-of-words, n-grams, and phrase detection Frequency-based filtering and feature scaling for eliminating uninformative

features Encoding techniques of categorical variables, including feature hashing and bin-counting Model-based feature engineering with principal component analysis The concept of model stacking, using k-means as a featurization technique Image feature extraction with manual and deep-learning techniques

Regression Modeling Strategies

Predictive Analytics For Dummies

Quickly and Easily Write Dynamic Documents Suitable for both beginners and advanced users, Dynamic Documents with R and knitr, Second Edition makes writing statistical reports easier by integrating computing directly with reporting. Reports range from homework, projects, exams, books, blogs, and web pages to virtually any documents related to statistical graphics, computing, and data analysis. The book covers basic applications for beginners while guiding power users in understanding the extensibility of the knitr package. New to the Second Edition A new chapter that introduces R Markdown v2 Changes that reflect improvements in the knitr package New sections on generating tables, defining custom printing methods for objects in code chunks, the C/Fortran engines, the Stan engine, running engines in a persistent session, and starting a local server to

serve dynamic documents Boost Your Productivity in Statistical Report Writing and Make Your Scientific Computing with R Reproducible Like its highly praised predecessor, this edition shows you how to improve your efficiency in writing reports. The book takes you from program output to publication-quality reports, helping you fine-tune every aspect of your report.

Bayesian Networks

Applied Predictive Modeling covers the overall predictive modeling process, beginning with the crucial steps of data preprocessing, data splitting and foundations of model tuning. The text then provides intuitive explanations of numerous common and modern regression and classification techniques, always with an emphasis on illustrating and solving real data problems. The text illustrates all parts of the modeling process through many hands-on, real-life examples, and every chapter contains extensive R code for each step of the process. This multi-purpose text can be used as an introduction to predictive models and the overall modeling process, a practitioner's reference handbook, or as a text for advanced undergraduate or graduate level predictive modeling courses. To that end, each chapter contains problem sets to help solidify the covered concepts and uses data available in the book's R package. This text is intended for a broad audience as both an introduction to predictive models as well as a guide to applying them. Non-mathematical readers will appreciate the intuitive explanations of the techniques

while an emphasis on problem-solving with real data across a wide variety of applications will aid practitioners who wish to extend their expertise. Readers should have knowledge of basic statistical ideas, such as correlation and linear regression analysis. While the text is biased against complex equations, a mathematical background is needed for advanced topics.

Statistical Learning from a Regression Perspective

Build machine learning algorithms, prepare data, and dig deep into data prediction techniques with R About This Book Harness the power of R for statistical computing and data science Explore, forecast, and classify data with R Use R to apply common machine learning algorithms to real-world scenarios Who This Book Is For Perhaps you already know a bit about machine learning but have never used R, or perhaps you know a little R but are new to machine learning. In either case, this book will get you up and running quickly. It would be helpful to have a bit of familiarity with basic programming concepts, but no prior experience is required. What You Will Learn Harness the power of R to build common machine learning algorithms with real-world data science applications Get to grips with techniques in R to clean and prepare your data for analysis and visualize your results Discover the different types of machine learning models and learn what is best to meet your data needs and solve data analysis problems Classify your data with Bayesian and nearest neighbour methods Predict values using R to build decision trees, rules,

and support vector machines Forecast numeric values with linear regression and model your data with neural networks Evaluate and improve the performance of machine learning models Learn specialized machine learning techniques for text mining, social network data, and big data In Detail Machine learning, at its core, is concerned with transforming data into actionable knowledge. This makes machine learning well suited to the present-day era of big data. Given the growing prominence of R's cross-platform, zero-cost statistical programming environment, there has never been a better time to start applying machine learning to your data. Machine learning with R offers a powerful set of methods to quickly and easily gain insight from your data to both, veterans and beginners in data analytics. Want to turn your data into actionable knowledge, predict outcomes that make real impact, and have constantly developing insights? R gives you access to all the power you need to master exceptional machine learning techniques. The second edition of Machine Learning with R provides you with an introduction to the essential skills required in data science. Without shying away from technical theory, it is written to provide focused and practical knowledge to get you building algorithms and crunching your data, with minimal previous experience. With this book, you'll discover all the analytical tools you need to gain insights from complex data and learn to to choose the correct algorithm for your specific needs. Through full engagement with the sort of real-world problems data-wranglers face, you'll learn to apply machine learning methods to deal with common tasks, including classification, prediction, forecasting, market analysis, and clustering. Transform

the way you think about data; discover machine learning with R. Style and approach How can we use machine learning to transform data into action? This book uses a series of simple steps to show you. Using practical examples, the book illustrates how to prepare data for analysis, choose a machine learning method, and measure its success.

Applied Predictive Analytics

If you're an experienced programmer interested in crunching data, this book will get you started with machine learning—a toolkit of algorithms that enables computers to train themselves to automate useful tasks. Authors Drew Conway and John Myles White help you understand machine learning and statistics tools through a series of hands-on case studies, instead of a traditional math-heavy presentation. Each chapter focuses on a specific problem in machine learning, such as classification, prediction, optimization, and recommendation. Using the R programming language, you'll learn how to analyze sample datasets and write simple machine learning algorithms. Machine Learning for Hackers is ideal for programmers from any background, including business, government, and academic research. Develop a naïve Bayesian classifier to determine if an email is spam, based only on its text Use linear regression to predict the number of page views for the top 1,000 websites Learn optimization techniques by attempting to break a simple letter cipher Compare and contrast U.S. Senators statistically,

based on their voting records Build a "whom to follow" recommendation system from Twitter data

Hands-On Machine Learning with R

This book focuses on statistical methods for the analysis of discrete failure times. Failure time analysis is one of the most important fields in statistical research, with applications affecting a wide range of disciplines, in particular, demography, econometrics, epidemiology and clinical research. Although there are a large variety of statistical methods for failure time analysis, many techniques are designed for failure times that are measured on a continuous scale. In empirical studies, however, failure times are often discrete, either because they have been measured in intervals (e.g., quarterly or yearly) or because they have been rounded or grouped. The book covers well-established methods like life-table analysis and discrete hazard regression models, but also introduces state-of-the art techniques for model evaluation, nonparametric estimation and variable selection. Throughout, the methods are illustrated by real life applications, and relationships to survival analysis in continuous time are explained. Each section includes a set of exercises on the respective topics. Various functions and tools for the analysis of discrete survival data are collected in the R package discSurv that accompanies the book.

Applied Predictive Modeling

This book provides a general and comprehensible overview of imbalanced learning. It contains a formal description of a problem, and focuses on its main features, and the most relevant proposed solutions. Additionally, it considers the different scenarios in Data Science for which the imbalanced classification can create a real challenge. This book stresses the gap with standard classification tasks by reviewing the case studies and ad-hoc performance metrics that are applied in this area. It also covers the different approaches that have been traditionally applied to address the binary skewed class distribution. Specifically, it reviews cost-sensitive learning, data-level preprocessing methods and algorithm-level solutions, taking also into account those ensemble-learning solutions that embed any of the former alternatives. Furthermore, it focuses on the extension of the problem for multi-class problems, where the former classical methods are no longer to be applied in a straightforward way. This book also focuses on the data intrinsic characteristics that are the main causes which, added to the uneven class distribution, truly hinders the performance of classification algorithms in this scenario. Then, some notes on data reduction are provided in order to understand the advantages related to the use of this type of approaches. Finally this book introduces some novel areas of study that are gathering a deeper attention on the imbalanced data issue. Specifically, it considers the classification of data streams, non-classical classification problems, and the scalability related to Big Data. Examples of

software libraries and modules to address imbalanced classification are provided. This book is highly suitable for technical professionals, senior undergraduate and graduate students in the areas of data science, computer science and engineering. It will also be useful for scientists and researchers to gain insight on the current developments in this area of study, as well as future research directions.

Interpretable Machine Learning

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.

Hands-On Predictive Analytics with Python

Learn methods of data analysis and their application to real-world data sets This updated second edition serves as an introduction to data mining methods and models, including association rules, clustering, neural networks, logistic regression, and multivariate analysis. The authors apply a unified "white box" approach to data mining methods and models. This approach is designed to walk readers through the operations and nuances of the various methods, using small data sets, so readers can gain an insight into the inner workings of the method under review.

Chapters provide readers with hands-on analysis problems, representing an opportunity for readers to apply their newly-acquired data mining expertise to solving real problems using large, real-world data sets. Data Mining and Predictive Analytics: Offers comprehensive coverage of association rules, clustering, neural networks, logistic regression, multivariate analysis, and R statistical programming language Features over 750 chapter exercises, allowing readers to assess their understanding of the new material Provides a detailed case study that brings together the lessons learned in the book Includes access to the companion website, www.dataminingconsultant, with exclusive password-protected instructor content Data Mining and Predictive Analytics will appeal to computer science and statistic students, as well as students in MBA programs, and chief executives.

Learning from Imbalanced Data Sets

Now , a leader of Northwestern University's prestigious analytics program presents a fully-integrated treatment of both the business and academic elements of marketing applications in predictive analytics. Writing for both managers and students, Thomas W. Miller explains essential concepts, principles, and theory in the context of real-world applications. Building on Miller's pioneering program, Marketing Data Science thoroughly addresses segmentation, target marketing, brand and product positioning, new product development, choice modeling, recommender systems, pricing research, retail site selection, demand estimation,

sales forecasting, customer retention, and lifetime value analysis. Starting where Miller's widely-praised Modeling Techniques in Predictive Analytics left off, he integrates crucial information and insights that were previously segregated in texts on web analytics, network science, information technology, and programming. Coverage includes: The role of analytics in delivering effective messages on the web Understanding the web by understanding its hidden structures Being recognized on the web – and watching your own competitors Visualizing networks and understanding communities within them Measuring sentiment and making recommendations Leveraging key data science methods: databases/data preparation, classical/Bayesian statistics, regression/classification, machine learning, and text analytics Six complete case studies address exceptionally relevant issues such as: separating legitimate email from spam; identifying legally-relevant information for lawsuit discovery; gleaning insights from anonymous web surfing data, and more. This text's extensive set of web and network problems draw on rich public-domain data sources; many are accompanied by solutions in Python and/or R. Marketing Data Science will be an invaluable resource for all students, faculty, and professional marketers who want to use business analytics to improve marketing performance.

Predictive Analytics, Data Mining and Big Data

Statistical Inference via Data Science: A ModernDive into R and the Tidyverse

provides a pathway for learning about statistical inference using data science tools widely used in industry, academia, and government. It introduces the tidyverse suite of R packages, including the ggplot2 package for data visualization, and the dplyr package for data wrangling. After equipping readers with just enough of these data science tools to perform effective exploratory data analyses, the book covers traditional introductory statistics topics like confidence intervals, hypothesis testing, and multiple regression modeling, while focusing on visualization throughout. Features: ● Assumes minimal prerequisites, notably, no prior calculus nor coding experience ● Motivates theory using real-world data, including all domestic flights leaving New York City in 2013, the Gapminder project, and the data journalism website, FiveThirtyEight.com ● Centers on simulation-based approaches to statistical inference rather than mathematical formulas ● Uses the infer package for "tidy" and transparent statistical inference to construct confidence intervals and conduct hypothesis tests via the bootstrap and permutation methods ● Provides all code and output embedded directly in the text; also available in the online version at moderndive.com This book is intended for individuals who would like to simultaneously start developing their data science toolbox and start learning about the inferential and modeling tools used in much of modern-day research. The book can be used in methods and data science courses and first courses in statistics, at both the undergraduate and graduate levels.

Sentence logic

Machine learning methods are now an important tool for scientists, researchers, engineers and students in a wide range of areas. This book is written for people who want to adopt and use the main tools of machine learning, but aren't necessarily going to want to be machine learning researchers. Intended for students in final year undergraduate or first year graduate computer science programs in machine learning, this textbook is a machine learning toolkit. Applied Machine Learning covers many topics for people who want to use machine learning processes to get things done, with a strong emphasis on using existing tools and packages, rather than writing one's own code. A companion to the author's Probability and Statistics for Computer Science, this book picks up where the earlier book left off (but also supplies a summary of probability that the reader can use). Emphasizing the usefulness of standard machinery from applied statistics, this textbook gives an overview of the major applied areas in learning, including coverage of:• classification using standard machinery (naive bayes; nearest neighbor; SVM)• clustering and vector quantization (largely as in PSCS)• PCA (largely as in PSCS)• variants of PCA (NIPALS; latent semantic analysis; canonical correlation analysis)• linear regression (largely as in PSCS)• generalized linear models including logistic regression• model selection with Lasso, elasticnet• robustness and m-estimators• Markov chains and HMM's (largely as in PSCS)• EM in fairly gory detail; long experience teaching this suggests one detailed example is required, which students hate; but once they've been through that, the next one

is easy• simple graphical models (in the variational inference section)• classification with neural networks, with a particular emphasis onimage classification• autoencoding with neural networks• structure learning

An Introduction to Stochastic Dynamics

This in-depth guide provides managers with a solid understanding of data and data trends, the opportunities that it can offer to businesses, and the dangers of these technologies. Written in an accessible style, Steven Finlay provides a contextual roadmap for developing solutions that deliver benefits to organizations.

Data Mining and Predictive Analytics

R offers a free and open source environment that is perfect for both learning and deploying predictive modeling solutions in the real world. With its constantly growing community and plethora of packages, R offers the functionality to deal with a truly vast array of problems. This book is designed to be both a guide and a reference for moving beyond the basics of predictive modeling. The book begins with a dedicated chapter on the language of models and the predictive modeling process. Each subsequent chapter tackles a particular type of model, such as neural networks, and focuses on the three important questions of how the model

works, how to use R to train it, and how to measure and assess its performance using real world data sets. By the end of this book, you will have explored and tested the most popular modeling techniques in use on real world data sets and mastered a diverse range of techniques in predictive analytics.

Statistical Inference via Data Science: A ModernDive into R and the Tidyverse

Two-volume introduction to formal logic. Volume I presents sentence logic and Volume II covers predicate logic and metatheory. Features easy-to-understand explanations and graded exercises.

Machine Learning with R

An accessible introduction for applied mathematicians to concepts and techniques for describing, quantifying, and understanding dynamics under uncertainty.

Modeling to Inform Infectious Disease Control

"Mesmerizing & fascinating" —The Seattle Post-Intelligencer "The Freakonomics of big data." —Stein Kretsinger, founding executive of Advertising.com Award-winning

| Used by over 30 universities | Translated into 9 languages An introduction for everyone. In this rich, fascinating — surprisingly accessible — introduction, leading expert Eric Siegel reveals how predictive analytics (aka machine learning) works, and how it affects everyone every day. Rather than a "how to" for hands-on techies, the book serves lay readers and experts alike by covering new case studies and the latest state-of-the-art techniques. Prediction is booming. It reinvents industries and runs the world. Companies, governments, law enforcement, hospitals, and universities are seizing upon the power. These institutions predict whether you're going to click, buy, lie, or die. Why? For good reason: predicting human behavior combats risk, boosts sales, fortifies healthcare, streamlines manufacturing, conquers spam, optimizes social networks, toughens crime fighting, and wins elections. How? Prediction is powered by the world's most potent, flourishing unnatural resource: data. Accumulated in large part as the by-product of routine tasks, data is the unsalted, flavorless residue deposited en masse as organizations churn away. Surprise! This heap of refuse is a gold mine. Big data embodies an extraordinary wealth of experience from which to learn. Predictive analytics (aka machine learning) unleashes the power of data. With this technology, the computer literally learns from data how to predict the future behavior of individuals. Perfect prediction is not possible, but putting odds on the future drives millions of decisions more effectively, determining whom to call, mail, investigate, incarcerate, set up on a date, or medicate. In this lucid, captivating introduction — now in its Revised and Updated edition — former Columbia

University professor and Predictive Analytics World founder Eric Siegel reveals the power and perils of prediction: What type of mortgage risk Chase Bank predicted before the recession. Predicting which people will drop out of school, cancel a subscription, or get divorced before they even know it themselves. Why early retirement predicts a shorter life expectancy and vegetarians miss fewer flights. Five reasons why organizations predict death — including one health insurance company. How U.S. Bank and Obama for America calculated the way to most strongly persuade each individual. Why the NSA wants all your data: machine learning supercomputers to fight terrorism. How IBM's Watson computer used predictive modeling to answer questions and beat the human champs on TV's Jeopardy! How companies ascertain untold, private truths — how Target figures out you're pregnant and Hewlett-Packard deduces you're about to quit your job. How judges and parole boards rely on crime-predicting computers to decide how long convicts remain in prison. 182 examples from Airbnb, the BBC, Citibank, ConEd, Facebook, Ford, Google, the IRS, LinkedIn, Match.com, MTV, Netflix, PayPal, Pfizer, Spotify, Uber, UPS, Wikipedia, and more. How does predictive analytics work? This jam-packed book satisfies by demystifying the intriguing science under the hood. For future hands-on practitioners pursuing a career in the field, it sets a strong foundation, delivers the prerequisite knowledge, and whets your appetite for more. A truly omnipresent science, predictive analytics constantly affects our daily lives. Whether you are a consumer of it — or consumed by it — get a handle on the power of Predictive Analytics.

Learning Predictive Analytics with Python

A hands-on introduction to the principles of Bayesian modeling using WinBUGS Bayesian Modeling Using WinBUGS provides an easily accessible introduction to the use of WinBUGS programming techniques in a variety of Bayesian modeling settings. The author provides an accessible treatment of the topic, offering readers a smooth introduction to the principles of Bayesian modeling with detailed guidance on the practical implementation of key principles. The book begins with a basic introduction to Bayesian inference and the WinBUGS software and goes on to cover key topics, including: Markov Chain Monte Carlo algorithms in Bayesian inference Generalized linear models Bayesian hierarchical models Predictive distribution and model checking Bayesian model and variable evaluation Computational notes and screen captures illustrate the use of both WinBUGS as well as R software to apply the discussed techniques. Exercises at the end of each chapter allow readers to test their understanding of the presented concepts and all data sets and code are available on the book's related Web site. Requiring only a working knowledge of probability theory and statistics, Bayesian Modeling Using WinBUGS serves as an excellent book for courses on Bayesian statistics at the upper-undergraduate and graduate levels. It is also a valuable reference for researchers and practitioners in the fields of statistics, actuarial science, medicine, and the social sciences who use WinBUGS in their everyday work.

Statistical Modeling and Computation

The process of developing predictive models includes many stages. Most resources focus on the modeling algorithms but neglect other critical aspects of the modeling process. This book describes techniques for finding the best representations of predictors for modeling and for nding the best subset of predictors for improving model performance. A variety of example data sets are used to illustrate the techniques along with R programs for reproducing the results.

Marketing Data Science

Examine business problems and use a practical analytical approach to solve them by implementing predictive models and machine learning techniques using SAS and the R analytical language. This book is ideal for those who are well-versed in writing code and have a basic understanding of statistics, but have limited experience in implementing predictive models and machine learning techniques for analyzing real world data. The most challenging part of solving industrial business problems is the practical and hands-on knowledge of building and deploying advanced predictive models and machine learning algorithms. Applied Analytics through Case Studies Using SAS and R is your answer to solving these business problems by sharpening your analytical skills. What You'll Learn

Understand analytics and basic data concepts Use an analytical approach to solve Industrial business problems Build predictive model with machine learning techniques Create and apply analytical strategies Who This Book Is For Data scientists, developers, statisticians, engineers, and research students with a great theoretical understanding of data and statistics who would like to enhance their skills by getting practical exposure in data modeling.

Feature Engineering for Machine Learning

Effectively Assess Intervention Options for Controlling Infectious DiseasesOur experiences with the human immunodeficiency virus (HIV), severe acute respiratory syndrome (SARS), and Ebola virus disease (EVD) remind us of the continuing need to be vigilant against the emergence of new infectious diseases. Mathematical modeling is increasingly used i

ROMANCE ACTION & ADVENTURE MYSTERY & THRILLER BIOGRAPHIES & HISTORY CHILDREN'S YOUNG ADULT FANTASY HISTORICAL FICTION HORROR LITERARY FICTION NON-FICTION SCIENCE FICTION