Practical Statistics For Data Scientists 50 Essential Concepts

Practical Statistics for Data Scientists **Statistics for Data Scientists Doing Data Science** Foundations of Data Science Data Pipelines Pocket Reference *Data Mining for Business Analytics Science: 50 Essential Ideas* **Essential MATLAB for Scientists and Engineers Applied Wavelet Analysis with S-PLUS 30-Second Data Science Boom! 50 Fantastic Science Experiments to Try at Home with Your Kids (PB)** Effective Data Science Infrastructure Women in Science **SQL for Data Scientists Data Science from Scratch 50 Science Things to Make and Do** *Strengthening Forensic Science in the United States* 30-Second Forensic Science **Python Data Science Handbook** *Data Science Foundations Tools and Techniques Python for Data Analysis Marvels of Science R for Data Science* **Data Science** *Deep Learning from Scratch* **SCIENCE** *Unsettled* **Basic Science Methods for Clinical Researchers Data Science for Business How to Tell the Truth with Statistics** Foundations of Statistics for Data Scientists Introduction to Machine Learning with Python **Essentials of Glycobiology Deep Learning for the Life Sciences Becoming a Data Head** What Is Color? **The Science of Effective Mentorship in STEMM** *The Secret of the Island Learning Spark* 30-Second Theories

As recognized, adventure as competently as experience roughly lesson, amusement, as with ease as pact can be gotten by just checking out a ebook **Practical Statistics For Data Scientists 50 Essential Concepts** in addition to it is not directly done, you could undertake even more roughly speaking this life, all but the world.

We provide you this proper as well as simple exaggeration to acquire those all. We meet the expense of Practical Statistics For Data Scientists 50 Essential Concepts and numerous ebook collections from fictions to scientific research in any way. accompanied by them is this Practical Statistics For Data Scientists 50 Essential Concepts that can be your partner.

R for Data Science Dec 04 2020 Learn how to use R to turn raw data into insight, knowledge, and understanding. This book introduces you to R, RStudio, and the tidyverse, a collection of R packages designed to work together to make data science fast, fluent, and fun. Suitable for readers with no previous programming experience, R for Data Science is designed to get you doing data science as quickly as possible. Authors Hadley Wickham and Garrett Grolemund guide you through the steps of importing, wrangling, exploring, and modeling your data and communicating the results. You'll get a complete, big-picture understanding of the data science cycle, along with basic tools you need to manage the details. Each section of the book is paired with exercises to help you practice what you've learned along the way. You'll learn how to: Wrangle—transform your datasets into a form convenient for analysis Program—learn powerful R tools for solving data problems with greater clarity and ease Explore—examine your data, generate hypotheses, and quickly test them Model—provide a low-dimensional summary that captures true "signals" in your dataset Communicate—learn R Markdown for integrating prose, code, and results

Practical Statistics for Data Scientists Oct 26 2022 Statistical methods are a key part of data science, yet few data scientists have formal statistical training. Courses and books on basic statistics rarely cover the topic from a data science perspective. The second edition of this practical guide--now including examples in Python as well as R--explains how to apply various statistical methods to data science, tells you how to avoid their misuse, and gives you advice on what's important and what's not. Many data scientists use statistical methods but lack a deeper statistical perspective. If you're familiar with the R or Python programming languages, and have had some exposure to statistics but want to learn more, this quick reference bridges the gap in an accessible, readable format. With this updated edition, you'll dive into: Exploratory data analysis Data and sampling distributions Statistical experiments and significance testing Regression and prediction Classification Statistical machine learning Unsupervised learning

SCIENCE Sep 01 2020

Foundations of Statistics for Data Scientists Mar 27 2020 Foundations of Statistics for Data Scientists: With R and Python is designed as a textbook for a one- or two-term introduction to mathematical statistics for students training to become data scientists. It is an in-depth presentation of the topics

in statistical science with which any data scientist should be familiar, including probability distributions, descriptive and inferential statistical methods, and linear modeling. The book assumes knowledge of basic calculus, so the presentation can focus on "why it works" as well as "how to do it." Compared to traditional "mathematical statistics" textbooks, however, the book has less emphasis on probability theory and more emphasis on using software to implement statistical methods and to conduct simulations to illustrate key concepts. All statistical analyses in the book use R software, with an appendix showing the same analyses with Python. The book also introduces modern topics that do not normally appear in mathematical statistics texts but are highly relevant for data scientists, such as Bayesian inference, generalized linear models for non-normal responses (e.g., logistic regression and Poisson loglinear models), and regularized model fitting. The nearly 500 exercises are grouped into "Data Analysis and Applications" and "Methods and Concepts." Appendices introduce R and Python and contain solutions for odd-numbered exercises. The book's website has expanded R, Python, and Matlab appendices and all data sets from the examples and exercises.

Boom! 50 Fantastic Science Experiments to Try at Home with Your Kids (PB) Dec 16 2021 This amazing book from the famous Naked Scientists offers a fun way to introduce science to kids, with 50 simple experiments that produce spectacular results. Want to know how to create fireworks from a bag of chips? Turn rice into quicksand? Generate a cloud in a soda bottle? How about build a toaster-powered hot air balloon, or work out the speed of light using margarine and a microwave? The results will amuse, astound, and educate in equal measure, whether you're 8 or 80. Most of these activities can be performed with commonplace materials that are probably lying around the house. Concise scientific explanations are included on how and why the experiments actually work. Each activity is straightforward and manageable, yet impressive enough to get anyone interested in science. So whether it's racing jelly jars, making a bowl invisible, or instantly freezing soda before your eyes—with the Naked Scientists' help, you'll never have a dull rainy day again!

Data Science from Scratch Aug 12 2021 Data science libraries, frameworks, modules, and toolkits are great for doing data science, but they're also a good way to dive into the discipline without actually understanding data science. In this book, you'll learn how many of the most fundamental data science tools and algorithms work by implementing them from scratch. If you have an

aptitude for mathematics and some programming skills, author Joel Grus will help you get comfortable with the math and statistics at the core of data science, and with hacking skills you need to get started as a data scientist. Today's messy glut of data holds answers to questions no one's even thought to ask. This book provides you with the know-how to dig those answers out. Get a crash course in Python Learn the basics of linear algebra, statistics, and probability—and understand how and when they're used in data science Collect, explore, clean, munge, and manipulate data Dive into the fundamentals of machine learning Implement models such as k-nearest Neighbors, Naive Bayes, linear and logistic regression, decision trees, neural networks, and clustering Explore recommender systems, natural language processing, network analysis, MapReduce, and databases

The Science of Effective Mentorship in STEMM Sep 20 2019 Mentorship is a catalyst capable of unleashing one's potential for discovery, curiosity, and participation in STEMM and subsequently improving the training environment in which that STEMM potential is fostered. Mentoring relationships provide developmental spaces in which students' STEMM skills are honed and pathways into STEMM fields can be discovered. Because mentorship can be so influential in shaping the future STEMM workforce, its occurrence should not be left to chance or idiosyncratic implementation. There is a gap between what we know about effective mentoring and how it is practiced in higher education. The Science of Effective Mentorship in STEMM studies mentoring programs and practices at the undergraduate and graduate levels. It explores the importance of mentorship, the science of mentoring relationships, mentorship of underrepresented students in STEMM, mentorship structures and behaviors, and institutional cultures that support mentorship. This report and its complementary interactive guide present insights on effective programs and practices that can be adopted and adapted by institutions, departments, and individual faculty members.

Python Data Science Handbook Apr 08 2021 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

Essentials of Glycobiology Jan 25 2020 Sugar chains (glycans) are often attached to proteins and lipids and have multiple roles in the organization and function of all organisms. "Essentials of Glycobiology" describes their biogenesis and function and offers a useful gateway to the understanding of glycans.

Data Mining for Business Analytics May 21 2022 An applied approach to data mining and predictive analytics with clear exposition, hands-on exercises, and real-life case studies. Readers will work with all of the standard data mining methods using the Microsoft® Office Excel® add-in XLMiner® to develop predictive models and learn how to obtain business value from Big Data. Featuring updated topical coverage on text mining, social network analysis, collaborative filtering, ensemble methods, uplift modeling and more, the Third Edition also includes: Real-world examples to build a theoretical and practical understanding of key data mining methods End-of-chapter exercises that help readers better understand the presented material Data-rich case studies to illustrate various applications of data mining techniques Completely new chapters on social network analysis and text mining A companion site with additional data sets, instructors material that include solutions to exercises and case studies, and Microsoft PowerPoint® slides https://www.dataminingbook.com Free 140-day license to use XLMiner for Education software Data Mining for Business Analytics: Concepts, Techniques, and Applications in XLMiner®, Third Edition is an ideal textbook for upper-undergraduate and graduate-level courses as well as professional programs on data mining, predictive modeling, and Big Data analytics. The new edition is also a unique reference for analysts, researchers, and practitioners working with predictive analytics in the fields of business, finance, marketing, computer science, and information technology. Praise for the Second Edition "…full of vivid and thought-provoking anecdotes... needs to be read by anyone with a serious interest in research and marketing."–

Research Magazine "Shmueli et al. have done a wonderful job in presenting the field of data mining - a welcome addition to the literature." – ComputingReviews.com "Excellent choice for business analysts...The book is a perfect fit for its intended audience." – Keith McCormick, Consultant and Author of SPSS Statistics For Dummies, Third Edition and SPSS Statistics for Data Analysis and Visualization Galit Shmueli, PhD, is Distinguished Professor at National Tsing Hua University's Institute of Service Science. She has designed and instructed data mining courses since 2004 at University of Maryland, Statistics.com, The Indian School of Business, and National Tsing Hua University, Taiwan. Professor Shmueli is known for her research and teaching in business analytics, with a focus on statistical and data mining methods in information systems and healthcare. She has authored over 70 journal articles, books, textbooks and book chapters. Peter C. Bruce is President and Founder of the Institute for Statistics Education at www.statistics.com. He has written multiple journal articles and is the developer of Resampling Stats software. He is the author of Introductory Statistics and Analytics: A Resampling Perspective, also published by Wiley. Nitin R. Patel, PhD, is Chairman and cofounder of Cytel, Inc., based in Cambridge, Massachusetts. A Fellow of the American Statistical Association, Dr. Patel has also served as a Visiting Professor at the Massachusetts Institute of Technology and at Harvard University. He is a Fellow of the Computer Society of India and was a professor at the Indian Institute of Management, Ahmedabad for 15 years.

SQL for Data Scientists Sep 13 2021 Jump-start your career as a data scientist—learn to develop datasets for exploration, analysis, and machine learning SQL for Data Scientists: A Beginner's Guide for Building Datasets for Analysis is a resource that's dedicated to the Structured Query Language (SQL) and dataset design skills that data scientists use most. Aspiring data scientists will learn how to how to construct datasets for exploration, analysis, and machine learning. You can also discover how to approach query design and develop SQL code to extract data insights while avoiding common pitfalls. You may be one of many people who are entering the field of Data Science from a range of professions and educational backgrounds, such as business analytics, social science, physics, economics, and computer science. Like many of them, you may have conducted analyses using spreadsheets as data sources, but never retrieved and engineered datasets from a relational database using SQL, which is a programming language designed for managing databases and extracting data. This guide for data

scientists differs from other instructional guides on the subject. It doesn't cover SQL broadly. Instead, you'll learn the subset of SQL skills that data analysts and data scientists use frequently. You'll also gain practical advice and direction on "how to think about constructing your dataset." Gain an understanding of relational database structure, query design, and SQL syntax Develop queries to construct datasets for use in applications like interactive reports and machine learning algorithms Review strategies and approaches so you can design analytical datasets Practice your techniques with the provided database and SQL code In this book, author Renee Teate shares knowledge gained during a 15-year career working with data, in roles ranging from database developer to data analyst to data scientist. She guides you through SQL code and dataset design concepts from an industry practitioner's perspective, moving your data scientist career forward!

Statistics for Data Scientists Sep 25 2022 A key component of data science is statistics and machine learning, but only a small proportion of data scientists are actually trained as statisticians. This concise guide illustrates how to apply statistical concepts essential to data science, with advice on how to avoid their misuse. Many courses and books teach basic statistics, but rarely from a data science perspective. And while many data science resources incorporate statistical methods, they typically lack a deep statistical perspective. This quick reference book bridges that gap in an accessible, readable format.

Python for Data Analysis Feb 06 2021 Get complete instructions for manipulating, processing, cleaning, and crunching datasets in Python. Updated for Python 3.6, the second edition of this hands-on guide is packed with practical case studies that show you how to solve a broad set of data analysis problems effectively. You'll learn the latest versions of pandas, NumPy, IPython, and Jupyter in the process. Written by Wes McKinney, the creator of the Python pandas project, this book is a practical, modern introduction to data science tools in Python. It's ideal for analysts new to Python and for Python programmers new to data science and scientific computing. Data files and related material are available on GitHub. Use the IPython shell and Jupyter notebook for exploratory computing Learn basic and advanced features in NumPy (Numerical Python) Get started with data analysis tools in the pandas library Use flexible tools to load, clean, transform, merge, and reshape data Create informative visualizations with matplotlib Apply the pandas groupby facility to slice, dice, and summarize datasets Analyze and manipulate regular and irregular time series data Learn

how to solve real-world data analysis problems with thorough, detailed examples

Basic Science Methods for Clinical Researchers Jun 29 2020 Basic Science Methods for Clinical Researchers addresses the specific challenges faced by clinicians without a conventional science background. The aim of the book is to introduce the reader to core experimental methods commonly used to answer questions in basic science research and to outline their relative strengths and limitations in generating conclusive data. This book will be a vital companion for clinicians undertaking laboratory-based science. It will support clinicians in the pursuit of their academic interests and in making an original contribution to their chosen field. In doing so, it will facilitate the development of tomorrow's clinician scientists and future leaders in discovery science. Serves as a helpful guide for clinical researchers who lack a conventional science background Organized around research themes pertaining to key biological molecules, from genes, to proteins, cells, and model organisms Features protocols, techniques for troubleshooting common problems, and an explanation of the advantages and limitations of a technique in generating conclusive data Appendices provide resources for practical research methodology, including legal frameworks for using stem cells and animals in the laboratory, ethical considerations, and good laboratory practice (GLP)

Strengthening Forensic Science in the United States Jun 10 2021 Scores of talented and dedicated people serve the forensic science community, performing vitally important work. However, they are often constrained by lack of adequate resources, sound policies, and national support. It is clear that change and advancements, both systematic and scientific, are needed in a number of forensic science disciplines to ensure the reliability of work, establish enforceable standards, and promote best practices with consistent application. Strengthening Forensic Science in the United States: A Path Forward provides a detailed plan for addressing these needs and suggests the creation of a new government entity, the National Institute of Forensic Science, to establish and enforce standards within the forensic science community. The benefits of improving and regulating the forensic science disciplines are clear: assisting law enforcement officials, enhancing homeland security, and reducing the risk of wrongful conviction and exoneration. Strengthening Forensic Science in the United States gives a full account of what is needed to advance the forensic science disciplines, including upgrading of systems and organizational structures, better training,

widespread adoption of uniform and enforceable best practices, and mandatory certification and accreditation programs. While this book provides an essential call-to-action for congress and policy makers, it also serves as a vital tool for law enforcement agencies, criminal prosecutors and attorneys, and forensic science educators.

Data Science for Business May 29 2020 Written by renowned data science experts Foster Provost and Tom Fawcett, Data Science for Business introduces the fundamental principles of data science, and walks you through the "data-analytic thinking" necessary for extracting useful knowledge and business value from the data you collect. This guide also helps you understand the many data-mining techniques in use today. Based on an MBA course Provost has taught at New York University over the past ten years, Data Science for Business provides examples of real-world business problems to illustrate these principles. You'll not only learn how to improve communication between business stakeholders and data scientists, but also how participate intelligently in your company's data science projects. You'll also discover how to think data-analytically, and fully appreciate how data science methods can support business decision-making. Understand how data science fits in your organization—and how you can use it for competitive advantage Treat data as a business asset that requires careful investment if you're to gain real value Approach business problems data-analytically, using the data-mining process to gather good data in the most appropriate way Learn general concepts for actually extracting knowledge from data Apply data science principles when interviewing data science job candidates

Women in Science Oct 14 2021 A gloriously illustrated celebration of trailblazing women. Women in Science highlights the contributions of fifty notable women to the STEM fields of science, technology, engineering and mathematics, from both the ancient and modern worlds. The book also contains fascinating infographics and an illustrated scientific glossary. The extraordinary women profiled include well-known figures like the physicist and chemist Marie Curie, as well as lesser-known pioneers such as Katherine Johnson, the African-American mathematician who calculated the trajectory of the Apollo 11 mission to the moon. Women in Science celebrates the achievements of the intrepid women who have paved the way for the next generation of female engineers, biologists, mathematicians, doctors, astronauts, physicists and beyond ...

Unsettled Jul 31 2020 "Unsettled is a remarkable book—probably the best book on climate change for the intelligent layperson—that achieves the feat

of conveying complex information clearly and in depth." —Claremont Review of Books "Surging sea levels are inundating the coasts." "Hurricanes and tornadoes are becoming fiercer and more frequent." "Climate change will be an economic disaster." You've heard all this presented as fact. But according to science, all of these statements are profoundly misleading. When it comes to climate change, the media, politicians, and other prominent voices have declared that "the science is settled." In reality, the long game of telephone from research to reports to the popular media is corrupted by misunderstanding and misinformation. Core questions—about the way the climate is responding to our influence, and what the impacts will be—remain largely unanswered. The climate is changing, but the why and how aren't as clear as you've probably been led to believe. Now, one of America's most distinguished scientists is clearing away the fog to explain what science really says (and doesn't say) about our changing climate. In Unsettled: What Climate Science Tells Us, What It Doesn't, and Why It Matters, Steven Koonin draws upon his decades of experience—including as a top science advisor to the Obama administration—to provide up-to-date insights and expert perspective free from political agendas. Fascinating, clear-headed, and full of surprises, this book gives readers the tools to both understand the climate issue and be savvier consumers of science media in general. Koonin takes readers behind the headlines to the more nuanced science itself, showing us where it comes from and guiding us through the implications of the evidence. He dispels popular myths and unveils little-known truths: despite a dramatic rise in greenhouse gas emissions, global temperatures actually decreased from 1940 to 1970. What's more, the models we use to predict the future aren't able to accurately describe the climate of the past, suggesting they are deeply flawed. Koonin also tackles society's response to a changing climate, using data-driven analysis to explain why many proposed "solutions" would be ineffective, and discussing how alternatives like adaptation and, if necessary, geoengineering will ensure humanity continues to prosper. Unsettled is a reality check buoyed by hope, offering the truth about climate science that you aren't getting elsewhere—what we know, what we don't, and what it all means for our future.

Effective Data Science Infrastructure Nov 15 2021 Simplify data science infrastructure to give data scientists an efficient path from prototype to production. In Effective Data Science Infrastructure you will learn how to: Design data science infrastructure that boosts productivity Handle compute and orchestration in the cloud Deploy machine learning to production

Monitor and manage performance and results Combine cloud-based tools into a cohesive data science environment Develop reproducible data science projects using Metaflow, Conda, and Docker Architect complex applications for multiple teams and large datasets Customize and grow data science infrastructure Effective Data Science Infrastructure: How to make data scientists more productive is a hands-on guide to assembling infrastructure for data science and machine learning applications. It reveals the processes used at Netflix and other data-driven companies to manage their cutting edge data infrastructure. In it, you'll master scalable techniques for data storage, computation, experiment tracking, and orchestration that are relevant to companies of all shapes and sizes. You'll learn how you can make data scientists more productive with your existing cloud infrastructure, a stack of open source software, and idiomatic Python. The author is donating proceeds from this book to charities that support women and underrepresented groups in data science. About the technology Growing data science projects from prototype to production requires reliable infrastructure. Using the powerful new techniques and tooling in this book, you can stand up an infrastructure stack that will scale with any organization, from startups to the largest enterprises. About the book Effective Data Science Infrastructure teaches you to build data pipelines and project workflows that will supercharge data scientists and their projects. Based on state-of-the-art tools and concepts that power data operations of Netflix, this book introduces a customizable cloud-based approach to model development and MLOps that you can easily adapt to your company's specific needs. As you roll out these practical processes, your teams will produce better and faster results when applying data science and machine learning to a wide array of business problems. What's inside Handle compute and orchestration in the cloud Combine cloud-based tools into a cohesive data science environment Develop reproducible data science projects using Metaflow, AWS, and the Python data ecosystem Architect complex applications that require large datasets and models, and a team of data scientists About the reader For infrastructure engineers and engineering-minded data scientists who are familiar with Python. About the author At Netflix, Ville Tuulos designed and built Metaflow, a full-stack framework for data science. Currently, he is the CEO of a startup focusing on data science infrastructure. Table of Contents 1 Introducing data science infrastructure 2 The toolchain of data science 3 Introducing Metaflow 4 Scaling with the compute layer 5 Practicing scalability and performance 6 Going to production 7 Processing data 8 Using and operating models 9 Machine learning with the

full stack

30-Second Data Science Jan 17 2022 30-Second Data Science covers 50 key elements in the field of data-gathering from basic to ethics, explained clearly and concisely.

Marvels of Science Jan 05 2021 Focusing on the characters, events, and moments of genius that comprise the story of science, these 50 short reads are ideal for both read-alouds and reading assignments. The tales take just minutes to read but amply illustrate scientific principles and the evolution of science through history. Discussion questions and additional references correlate each story with elements of the science curriculum and provide direction for students to pursue their own discoveries. Stories are cross-indexed by year of occurrence and by scientist.

What Is Color? Oct 22 2019 A comprehensive illustrated exploration of the fascinating science of color Arielle and Joann Eckstut, authors of The Secret Language of Color, offer a thorough, readable, and highly visual exploration of the science of color. Organized by 50 of the most essential questions about color across a variety of fields—physics, chemistry, biology, technology, and psychology—this book examines how and why we see color; how color relates to light; what the real primary colors are; how biology, language, and culture affect the colors that we see; and much more. Full of clear and elegant infographics, What Is Color? is a must-have for artists and designers, scientists, students, and decorators, and anyone else whose work or play involves color.

Data Science Nov 03 2020 Learn the basics of Data Science through an easy to understand conceptual framework and immediately practice using RapidMiner platform. Whether you are brand new to data science or working on your tenth project, this book will show you how to analyze data, uncover hidden patterns and relationships to aid important decisions and predictions. Data Science has become an essential tool to extract value from data for any organization that collects, stores and processes data as part of its operations. This book is ideal for business users, data analysts, business analysts, engineers, and analytics professionals and for anyone who works with data. You'll be able to: Gain the necessary knowledge of different data science techniques to extract value from data. Master the concepts and inner workings of 30 commonly used powerful data science algorithms. Implement step-by-step data science process using using RapidMiner, an open source GUI based data science platform Data Science techniques covered: Exploratory data analysis, Visualization, Decision trees, Rule induction, k-

nearest neighbors, Naïve Bayesian classifiers, Artificial neural networks, Deep learning, Support vector machines, Ensemble models, Random forests, Regression, Recommendation engines, Association analysis, K-Means and Density based clustering, Self organizing maps, Text mining, Time series forecasting, Anomaly detection, Feature selection and more... Contains fully updated content on data science, including tactics on how to mine business data for information Presents simple explanations for over twenty powerful data science techniques Enables the practical use of data science algorithms without the need for programming Demonstrates processes with practical use cases Introduces each algorithm or technique and explains the workings of a data science algorithm in plain language Describes the commonly used setup options for the open source tool RapidMiner

Becoming a Data Head Nov 22 2019 "Turn yourself into a Data Head. You'll become a more valuable employee and make your organization more successful." Thomas H. Davenport, Research Fellow, Author of Competing on Analytics, Big Data @ Work, and The AI Advantage You've heard the hype around data—now get the facts. In Becoming a Data Head: How to Think, Speak, and Understand Data Science, Statistics, and Machine Learning, award-winning data scientists Alex Gutman and Jordan Goldmeier pull back the curtain on data science and give you the language and tools necessary to talk and think critically about it. You'll learn how to: Think statistically and understand the role variation plays in your life and decision making Speak intelligently and ask the right questions about the statistics and results you encounter in the workplace Understand what's really going on with machine learning, text analytics, deep learning, and artificial intelligence Avoid common pitfalls when working with and interpreting data Becoming a Data Head is a complete guide for data science in the workplace: covering everything from the personalities you'll work with to the math behind the algorithms. The authors have spent years in data trenches and sought to create a fun, approachable, and eminently readable book. Anyone can become a Data Head—an active participant in data science, statistics, and machine learning. Whether you're a business professional, engineer, executive, or aspiring data scientist, this book is for you.

Deep Learning from Scratch Oct 02 2020 With the resurgence of neural networks in the 2010s, deep learning has become essential for machine learning practitioners and even many software engineers. This book provides a comprehensive introduction for data scientists and software engineers with machine learning experience. You'll start with deep learning basics and move

quickly to the details of important advanced architectures, implementing everything from scratch along the way. Author Seth Weidman shows you how neural networks work using a first principles approach. You'll learn how to apply multilayer neural networks, convolutional neural networks, and recurrent neural networks from the ground up. With a thorough understanding of how neural networks work mathematically, computationally, and conceptually, you'll be set up for success on all future deep learning projects. This book provides: Extremely clear and thorough mental models—accompanied by working code examples and mathematical explanations—for understanding neural networks Methods for implementing multilayer neural networks from scratch, using an easy-to-understand object-oriented framework Working implementations and clear-cut explanations of convolutional and recurrent neural networks Implementation of these neural network concepts using the popular PyTorch framework

The Secret of the Island Aug 20 2019 Books for All Kinds of Readers. ReadHowYouWant offers the widest selection of on-demand, accessible format editions on the market today. Our 7 different sizes of EasyRead are optimized by increasing the font size and spacing between the words and the letters. We partner with leading publishers around the globe. Our goal is to have accessible editions simultaneously released with publishers' new books so that all readers can have access to the books they want to read.

How to Tell the Truth with Statistics Apr 27 2020 Statistics has played a leading role in our scientific understanding of the world for centuries, yet we are all familiar with the way statistical claims can be sensationalised, particularly in the media. In the age of big data, as data science becomes established as a discipline, a basic grasp of statistical literacy is more important than ever. In How to Tell the Truth with Statistics, David Spiegelhalter guides the reader through the essential principles we need in order to derive knowledge from data. Drawing on real world problems to introduce conceptual issues, he shows us how statistics can help us determine the luckiest passenger on the Titanic, whether serial killer Harold Shipman could have been caught earlier, and if screening for ovarian cancer is beneficial. How many trees are there on the planet? Do busier hospitals have higher survival rates? Why do old men have big ears? Spiegelhalter reveals the answers to these and many other questions - questions that can only be addressed using statistical science.

Doing Data Science Aug 24 2022 Now that people are aware that data can make the difference in an election or a business model, data science as an

occupation is gaining ground. But how can you get started working in a wide-ranging, interdisciplinary field that's so clouded in hype? This insightful book, based on Columbia University's Introduction to Data Science class, tells you what you need to know. In many of these chapter-long lectures, data scientists from companies such as Google, Microsoft, and eBay share new algorithms, methods, and models by presenting case studies and the code they use. If you're familiar with linear algebra, probability, and statistics, and have programming experience, this book is an ideal introduction to data science. Topics include: Statistical inference, exploratory data analysis, and the data science process Algorithms Spam filters, Naive Bayes, and data wrangling Logistic regression Financial modeling Recommendation engines and causality Data visualization Social networks and data journalism Data engineering, MapReduce, Pregel, and Hadoop Doing Data Science is collaboration between course instructor Rachel Schutt, Senior VP of Data Science at News Corp, and data science consultant Cathy O'Neil, a senior data scientist at Johnson Research Labs, who attended and blogged about the course.

Learning Spark Jul 19 2019 Data is bigger, arrives faster, and comes in a variety of formats—and it all needs to be processed at scale for analytics or machine learning. But how can you process such varied workloads efficiently? Enter Apache Spark. Updated to include Spark 3.0, this second edition shows data engineers and data scientists why structure and unification in Spark matters. Specifically, this book explains how to perform simple and complex data analytics and employ machine learning algorithms. Through step-by-step walk-throughs, code snippets, and notebooks, you'll be able to: Learn Python, SQL, Scala, or Java high-level Structured APIs Understand Spark operations and SQL Engine Inspect, tune, and debug Spark operations with Spark configurations and Spark UI Connect to data sources: JSON, Parquet, CSV, Avro, ORC, Hive, S3, or Kafka Perform analytics on batch and streaming data using Structured Streaming Build reliable data pipelines with open source Delta Lake and Spark Develop machine learning pipelines with MLlib and productionize models using MLflow

Science: 50 Essential Ideas Apr 20 2022 How did life emerge? What are the smallest elements of matter? How are planets formed? Over the centuries, brilliant men and women have sought to develop theories to answer the most compelling questions about the world around us. Through their amazing insights and conscientious efforts they helped to create the world we know today. In this beautifully illustrated book, Anne Rooney introduces you to the

fascinating world of science and its greatest practitioners. Ranging from evolutionary biology to quantum physics to chaos theory and featuring the ideas of such pivotal scientists as Isaac Newton, Charles Darwin, and Stephen Hawking, this essential guide will bring you up to speed on all of the world's most important scientific discoveries.

Foundations of Data Science Jul 23 2022 Covers mathematical and algorithmic foundations of data science: machine learning, high-dimensional geometry, and analysis of large networks.

30-Second Forensic Science May 09 2021 Humanity's most appalling crimes are solved by experts presenting painstakingly gathered evidence to the court of law. Investigators rely on physical, chemical and digital clues gathered at the scene of an incident to reconstruct beyond all reasonable doubt the events that occurred in order to bring criminals to justice. Enter the forensic team, tasked with providing objective recognition and identification and evaluating physical evidence (the clues) to support known or suspected circumstances. Far from the super-sleuths of fiction, the real-life masters of deduction occupy a world of dogged detection, analysing fingerprints or gait, identifying traces of toxins, drugs or explosives, matching digital data, performing anatomical dissection, disease diagnosis, facial reconstruction and environmental profiling.

Introduction to Machine Learning with Python Feb 24 2020 Many Python developers are curious about what machine learning is and how it can be concretely applied to solve issues faced in businesses handling medium to large amount of data. Machine Learning with Python teaches you the basics of machine learning and provides a thorough hands-on understanding of the subject.You'll learn important machine learning concepts and algorithms, when to use them, and how to use them. The book will cover a machine learning workflow: data preprocessing and working with data, training algorithms, evaluating results, and implementing those algorithms into a production-level system.

Essential MATLAB for Scientists and Engineers Mar 19 2022 Based on a teach-yourself approach, the fundamentals of MATLAB are illustrated throughout with many examples from a number of different scientific and engineering areas, such as simulation, population modelling, and numerical methods, as well as from business and everyday life. Some of the examples draw on first-year university level maths, but these are self-contained so that their omission will not detract from learning the principles of using MATLAB. This completely revised new edition is based on the latest version

of MATLAB. New chapters cover handle graphics, graphical user interfaces (GUIs), structures and cell arrays, and importing/exporting data. The chapter on numerical methods now includes a general GUI-driver ODE solver. * Maintains the easy informal style of the first edition * Teaches the basic principles of scientific programming with MATLAB as the vehicle * Covers the latest version of MATLAB

50 Science Things to Make and Do Jul 11 2021 This handy book contains 50 stimulating activities -- make your own foaming monsters, hanging crystals, kaleidoscopes and more. A fresh approach to the practical world of science, combining creative craft activities with the basics of physics, chemistry and biology. Each activity that is accompanied by illustrated, step-by-step instructions. A great introduction to science for young children.

Applied Wavelet Analysis with S-PLUS Feb 18 2022 Using a visual data analysis approach, wavelet concepts are explained in a way that is intuitive and easy to understand. Furthermore, in addition to wavelets, a whole range of related signal processing techniques such as wavelet packets, local cosine analysis, and matching pursuits are covered, and applications of wavelet analysis are illustrated -including nonparametric function estimation, digital image compression, and time-frequency signal analysis. This book and software package is intended for a broad range of data analysts, scientists, and engineers. While most textbooks on the subject presuppose advanced training in mathematics, this book merely requires that readers be familiar with calculus and linear algebra at the undergraduate level.

Deep Learning for the Life Sciences Dec 24 2019 Deep learning has already achieved remarkable results in many fields. Now it's making waves throughout the sciences broadly and the life sciences in particular. This practical book teaches developers and scientists how to use deep learning for genomics, chemistry, biophysics, microscopy, medical analysis, and other fields. Ideal for practicing developers and scientists ready to apply their skills to scientific applications such as biology, genetics, and drug discovery, this book introduces several deep network primitives. You'll follow a case study on the problem of designing new therapeutics that ties together physics, chemistry, biology, and medicine—an example that represents one of science's greatest challenges. Learn the basics of performing machine learning on molecular data Understand why deep learning is a powerful tool for genetics and genomics Apply deep learning to understand biophysical systems Get a brief introduction to machine learning with DeepChem Use deep learning to analyze microscopic images Analyze medical scans using

deep learning techniques Learn about variational autoencoders and generative adversarial networks Interpret what your model is doing and how it's working

Data Pipelines Pocket Reference Jun 22 2022 Data pipelines are the foundation for success in data analytics. Moving data from numerous diverse sources and transforming it to provide context is the difference between having data and actually gaining value from it. This pocket reference defines data pipelines and explains how they work in today's modern data stack. You'll learn common considerations and key decision points when implementing pipelines, such as batch versus streaming data ingestion and build versus buy. This book addresses the most common decisions made by data professionals and discusses foundational concepts that apply to open source frameworks, commercial products, and homegrown solutions. You'll learn: What a data pipeline is and how it works How data is moved and processed on modern data infrastructure, including cloud platforms Common tools and products used by data engineers to build pipelines How pipelines support analytics and reporting needs Considerations for pipeline maintenance, testing, and alerting

Data Science Foundations Tools and Techniques Mar 07 2021 The Foundational Hands-On Skills You Need to Dive into Data Science "Freeman and Ross have created the definitive resource for new and aspiring data scientists to learn foundational programming skills." -From the foreword by Jared Lander, series editor Using data science techniques, you can transform raw data into actionable insights for domains ranging from urban planning to precision medicine. Programming Skills for Data Science brings together all the foundational skills you need to get started, even if you have no programming or data science experience. Leading instructors Michael Freeman and Joel Ross guide you through installing and configuring the tools you need to solve professional-level data science problems, including the widely used R language and Git version-control system. They explain how to wrangle your data into a form where it can be easily used, analyzed, and visualized so others can see the patterns you've uncovered. Step by step, you'll master powerful R programming techniques and troubleshooting skills for probing data in new ways, and at larger scales. Freeman and Ross teach through practical examples and exercises that can be combined into complete data science projects. Everything's focused on real-world application, so you can quickly start analyzing your own data and getting answers you can act upon. Learn to Install your complete data science environment, including R

and RStudio Manage projects efficiently, from version tracking to documentation Host, manage, and collaborate on data science projects with GitHub Master R language fundamentals: syntax, programming concepts, and data structures Load, format, explore, and restructure data for successful analysis Interact with databases and web APIs Master key principles for visualizing data accurately and intuitively Produce engaging, interactive visualizations with ggplot and other R packages Transform analyses into sharable documents and sites with R Markdown Create interactive web data science applications with Shiny Collaborate smoothly as part of a data science team Register your book for convenient access to downloads, updates, and/or corrections as they become available. See inside book for details.

30-Second Theories Jun 17 2019 When it comes to big science, very few things are conclusively known. From Quantum Mechanics to Natural Selection, what we have instead are theories - ideas explain why things happen the way they do. We don't know for certain these are correct - no one ever saw the Big Bang - but with them we can paint beautiful, breathtaking pictures of everything from human behaviour to what the future may hold. Profiling the key scientists behind each theory, "30-Second Theories" presents each entry in a unique, eye-catching full-colour design, with thought-provoking extras and stylish illustrations. It is essential for anyone keen on expanding their mind with science's most thrilling ideas.