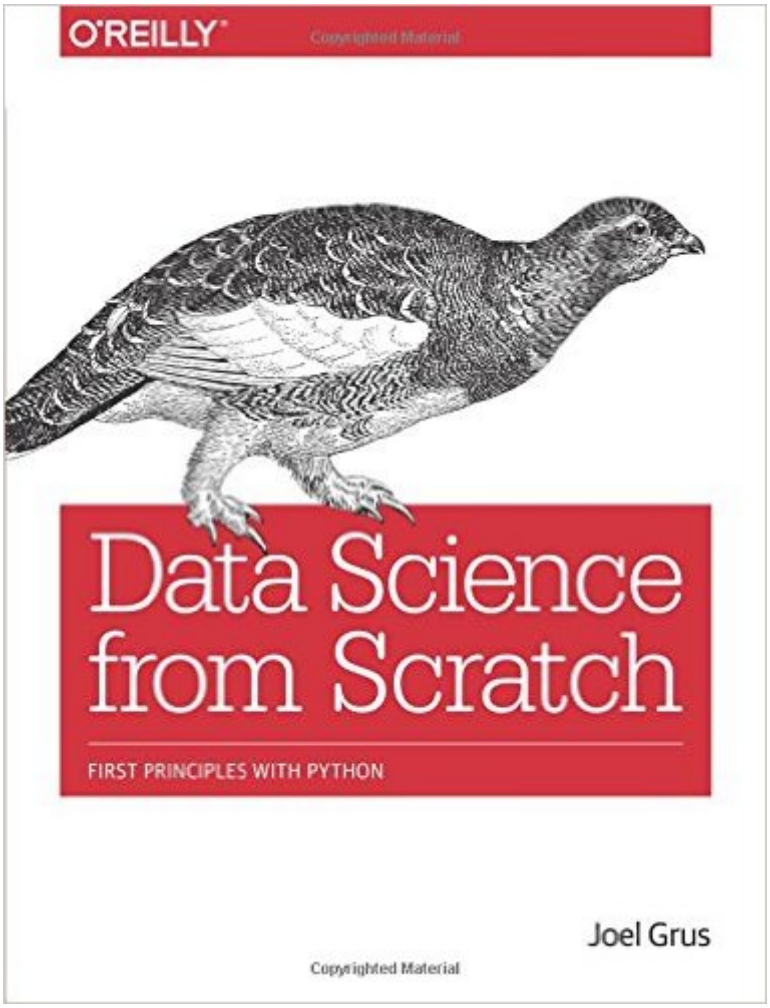
Data Science From Scratch: First Principles With Python



Synopsis

Data science libraries, frameworks, modules, and toolkits are great for doing data science, but theyâ ™re also a good way to dive into the discipline without actually understanding data science. In this book, youâ ™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. Todayâ ™s messy glut of data holds answers to questions no oneâ ™s even thought to ask. This book provides you with the know-how to dig those answers out.Get a crash course in PythonLearn the basics of linear algebra, statistics, and probabilityâ ”and understand how and when they’re used in data scienceCollect, explore, clean, munge, and manipulate dataDive into the fundamentals of machine learningImplement models such as k-nearest Neighbors, Naive Bayes, linear and logistic regression, decision trees, neural networks, and clusteringExplore recommender systems, natural language processing, network analysis, MapReduce, and databases

Book Information

Paperback: 330 pages

Publisher: O’Reilly Media; 1 edition (April 30, 2015)

Language: English

ISBN-10: 149190142X

ISBN-13: 978-1491901427

Product Dimensions: 6.9 x 0.7 x 9 inches

Shipping Weight: 1.1 pounds (View shipping rates and policies)

Average Customer Review: 4.2 out of 5 starsÂ Â See all reviewsÂ (64 customer reviews)

Best Sellers Rank: #5,105 in Books (See Top 100 in Books) #1 inÂ Books > Computers & Technology > Programming > Software Design, Testing & Engineering > Structured Design #2 inÂ Books > Textbooks > Computer Science > Database Storage & Design #3 inÂ Books > Computers & Technology > Databases & Big Data > Data Mining

Customer Reviews

This is a great book-- well written, easy to digest and informative. I’ve been in Data Mining and Statistical Analysis for a little over a decade now; I was looking for a book to share with my team to ensure we were all up-to-speed on some foundational concepts: this book is it. EDIT: I also forgot to mention, it has probably the best get-up-and-running in Python introduction I’ve seen (see, e.g.,

Chapter 2, ~20pp.)It’s the right size and correct coverage for the content and the author’s sense of humor (indeed, that of a data scientist) resonates with the audience.Solid introduction, even better review or brief explanation of commonly encountered topics.One of the best O’Reilly books I’ve read in a long time-- in fact, a technical book at the level I used to expect from O’Reilly.

The book is well-written and covers a wide range of topics related to data science and machine learning. Things I like about it:- The range of topics covered is wide, and includes (i) an intro/refresher for Python, (ii) statistics/probability, (iii) several ML techniques, (iv) data manipulation- For each topic, there is enough explanation of the underlying theory, as well as pointers to further reading- For each topic, the author builds up the code in simple steps, so it’s easy to follow along- Everything is explained very clearly; there is enough precision, without difficult formal language- The explanation of eigenvector centrality is awesome

This book emphasize excellent pedagogy and understandable Python code. The basis of all programming and mathematical algorithms is given with only an assumption of minimal prior programming and high school mathematics. The basics of Data Science including: 1. A 20 page Clear Introduction to Python 2, 2. An introduction to Linear Algebra (described by Python Functions) 3. A Similar Introduction to Practical Statistics.Like most scientific programmers who use Python the 2.6/2.7 branch is used throughout given the availability of appropriate libraries (like the Anaconda distribution). Tools for each type of algorithm are prototyped "from Scratch" in the author’s own exemplary code with references to the professional libraries in the final chapters. Math is for the most part taught from code rather than mathematical notation.Highly Recommended

This book is exactly what I needed to get started on data science.Prerequisites to benefit from the book :I really wanted a starter book, as despite the fact 30% of my university training was stats / probability, I’m very rusty on those topics. The books assume no prior maths knowledge beyond basics operations. I already code in Python so the refresher course was more a "which specific parts of Python are useful for Data science". I’ve learned a few things, but I assume that if you are entirely new to programming, this may be a bit tough (in that case I would recommend a python starter book such as Dive in Python 2 / Dive in Python 3).Note for Python 3 people : so far I had no major issues running the scripts with python 3, except a case of tuple unpacking that was easy to work around.Approach of the book :The book takes a example of the reader becoming chief data scientist in a dummy company, and the author provides the "business context" as needed (it’s web

oriented, but easy to grasp).What I love is that the book took the approach to build the tooling in plain python before pointing to the libraries : this serves as a support to explain the underlying maths, and it also avoid the "magic" effect of libraries, which makes it hard to solve issues as one don’t understand the underlying mechanism. Libraries are still used, but only after explaining the basics. This makes the knowledge more "portable" if you decide to use another library / language.Each chapter ends with a list of pointers to interesting online resources / libraries / etc, making it a good starter point to dig further.I highly recommend this book to anyone interested in learning data science, especially to those with rusty/limited math background.

This is among the handful of very best technical books I have ever read.As the "from Scratch" in the title implies, the objective of this book is to teach the fundamental ideas and techniques of data science from first (or nearly first) principles. After working through this book, you’ll be better able to meaningfully utilize the pre-packaged software (whether it’s Matlab, R, scikit-learn, or whatever) that you will use in "real life".And although the knowledge you’ll gain is largely independent of the programming language, you will as a bonus learn from the clear and elegant python code included. Every key topic, from probability, statistics, and other mathematical subjects, to machine learning and databases, is covered in a crystal clear manner.In summary, this book is the bee’s knees.

The book definitely provides the ability to do data analysis from scratch. However, I took that to mean it would provide me with the basic knowledge to do data science as its done in production. What you want find hardly anywhere in this book is a single discussion of how to use pandas, scikit-learn, numpy, matplotlib, or any of the other tools used by today’s data scientists.It does provide you with the basics of how the analysis is done and the math behind a bunch of machine learning models. Unfortunately there’s definitely a high expectation of math knowledge in this book, starting with at least Linear Algebra through Calculus.So, while I think this book had value, it was very different from what I was expecting. I think the target audience for this book is more an academic looking to apply their knowledge to a data science realm, and not so much a programmer wanting to learn how to actually implement machine learning or data analysis in code.

From?: Our First Talk About Birth (Just Enough) Programming ArcGIS with Python Cookbook - Second Edition The Definitive Guide to MongoDB: A complete guide to dealing with Big Data using MongoDB Big Data Fundamentals: Concepts, Drivers & Techniques (The Prentice Hall Service Technology Series from Thomas Erl) Building a Scalable Data Warehouse with Data Vault 2.0 Graph Databases: New Opportunities for Connected Data SQL: Beginner's Guide for Coding SQL (database programming, computer programming, how to program, sql for dummies, java, mysql, The Oracle, python, PHP, ... (HTML, Programming, Coding, CSS Book 7) Next Generation Databases: NoSQLand Big Data Next Generation Databases: NoSQL, NewSQL, and Big Data LEARN IN A DAY! DATA WAREHOUSING. Top Links and Resources for Learning Data Warehousing ONLINE and OFFLINE: Use these FREE and PAID resources to Learn Data Warehousing in little to no time My First Ramadan (My First Holiday) The Pilgrim Church: Being Some Account of the Continuance Through Succeeding Centuries of Churches Practising the Principles Taught and Exemplified in The New Testament Unity 5 From Zero to Proficiency (Foundations): A step-by-step guide to creating your first game with Unity. Pokemon the First Movie Ebook