Suggested Reads on Statistics & Mathematics for Data Science

Introduction

The selection process of data scientists at Google gives higher priority to candidates with strong background in statistics and mathematics. Not just Google, other top companies (Amazon, Airbnb, Uber etc.) in the world also prefer candidates with strong fundamentals rather than mere know-how in data science.

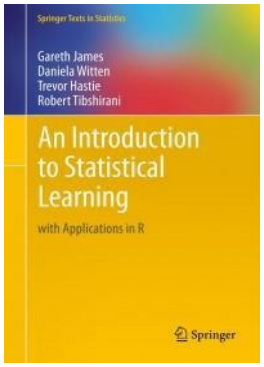
If you too aspire to work for such top companies in future, it is essential for you to develop a mathematical understanding of data science. Data science is simply the evolved version of statistics and mathematics, combined with programming and business logic. I've met many data scientists who struggle to explain predictive models statistically.

More than just deriving accuracy, understanding & interpreting every metric, calculation behind that accuracy is important. Remember, every single 'variable' has a story to tell. So, if not anything else, try to become a great story explorer!

In this document, we've compiled a list of books on statistics and mathematics. We understand, mathematics has no extreme. Hence, only those books are listed which will help you to connect with data science better.

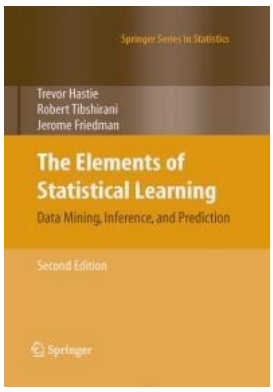
Statistics

Introduction to Statistical Learning



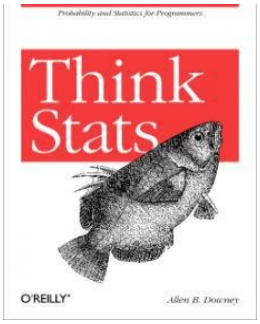
This is a highly recommended book for practicing data scientists. The focus of this books is kept on connecting statistics concept with machine learning. Hence, you'll learn about all popular supervised and unsupervised machine learning algorithms. R users will get an advantage, since the practical aspects of algorithms have been demonstrated using R. In addition to theory, this book also lay emphasis on using ML algorithms in real life setting. **Available:** Free Download

Elements of Statistical Learning



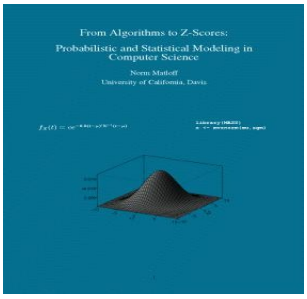
This book is an advanced level of previous book. It is written by Trevor Hastie and Rob Tibshirani, Professors at Stanford University. Their first book 'Introduction to Statistical Learning' uncover the basics of statistics and machine learning. This book, will introduce you to higher level algorithms such as Neural Networks, Bagging & Boosting, Kernel methods etc. The algorithms have been implemented in R programming. **Available:** Free Download

Think Stats



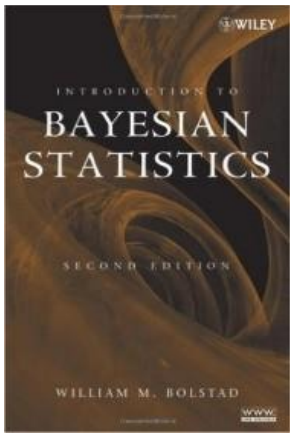
The author of this book is Alien B Downey. It is based on perform statistical analysis practically in Python. Hence, make sure you've got some basic knowledge of Python before buying this book. It focuses entirely on understanding real life influence of statistics using popular case studies. Since, stats and math are closely connected, it also has dedicated chapters on topic like bayesian estimation. **Available:** Buy from Amazon

From Algorithms to Z Scores



Did you know the about crucial role of statistics in programming? The author of this book is Norm Matloff, Professor, University of California. This book explains using probabilistic concepts and statistical measures in R. Again, a good practice source for R users. It teaches the art of dealing with probabilistic models and choosing the best one for final evaluation. It is a highly recommended book (specially for R users). **Available:** Free Download

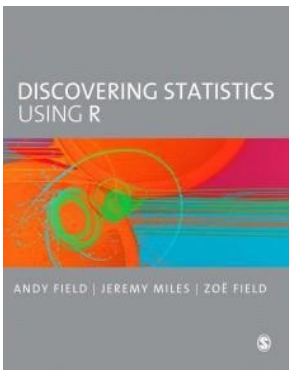
Introduction to Bayesian Statistics



This is a highly recommended book for freshers in data science. The author of this book is William M Bolstad. It's a must read for people who find mathematics boring. Having been written in a conversational style (rare to find math this way), this book is a great introductory resource on statistics. It begins with scientific methods of data gathering and end up delivering dedicated chapters on bayesian statistics.

Available: Free Download

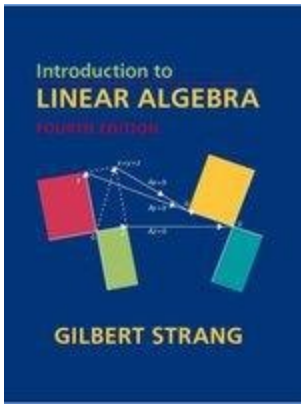
Discovering Statistics using R



This book is written by Andy Field, Jeremy Miles and Zoe Field. I would highly recommend this book to newbies in data science. To start with statistics, this book has a great content which goes in depth detail of its topics. Along with, the statistical concept are explained in conjunction with R which makes it even more useful. It offers a step by step understanding, with a parallel support of interesting practice examples. **Available:** Buy on Amazon

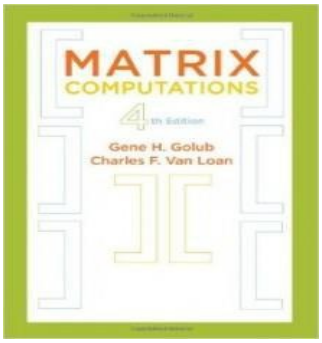
Mathematics

Introduction to Linear Algebra



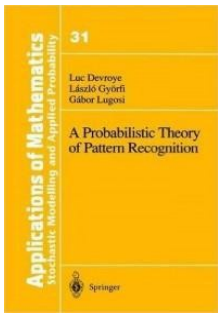
This is one of the most recommended book on Linear Algebra. The author of this book is Gilbert Strang, Professor, MIT. Gilbert unique way of delivering knowledge would give you the intuition and excitement to move forward after every chapter. This book will help you to build a strong mathematical foundation for machine learning. It enlists all the necessary chapters such as vectors, linear equations, determinants, eigenvalues, matrix factorization etc in great depth. **Available:** Buy on Amazon

Matrix Computation



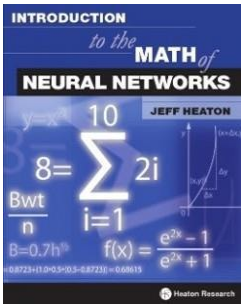
Matrix and Data frames are essential components of machine learning. The author of this book is Gene H Golub and Charles F Van Loan. This book provides a nice head start to students with concepts of matrix computations. The author covers most of the important topics such as gaussian elimination, matrix factorization, lancoz method, error analysis etc. Every chapter is supported by intuitive practice problems. The pseudo codes are available in Matlab. **Available:** Free Download

A Probabilistic Theory of Pattern Recognition



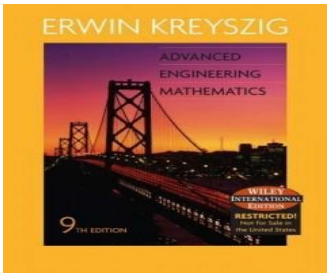
This is a complete resource to learn application of mathematics. This is a must read book for intermediate and advanced practitioners in machine learning. This book is written by Luc Devroye, Laszlo Gyorfi and Gabor Lugosi. It covers a wide range of topics varying from bayes error, linear discrimination to epsilon entropy & neural networks. It provides a convincing explanation to complex theorems with section wise practice problems. **Available:** Free Download

Introduction of Math of Neural Networks



If you have innate interest in learning about neural network, this should be your place to start. The author of this book is Jeff Heaton. The author has beautifully simplified the difficult concepts of neural networks. This book introduces you to basics of underlying maths in neural networks. It assumes reader has prior knowledge of algebra, calculus and programming. It demonstrates various mathematical tools which can be applied to neural networks. **Available:** Buy on Amazon

Advanced Engineering Mathematics



This is probably the most comprehensive book available on mathematics for machine learning users. The author of this book is Erwin Kreyszig. As a matter of fact, this book is highly recommended to college students as well. If you haven't been good at maths till now, follow this book religiously and you should surely see significant improvements in your math understanding. Along with derivations & practice example, this book has dedicated sections of calculus, algebra, probability etc. Definitely, a must read book for all levels of practitioners in data science. **Available:** Free Download

Cookbook on Probability and Statistics

This cookbook is must have in your digital bookshelf. This isn't exactly a text book you'd discover, but a quick digital guide on mathematical equations. The author of this book is Matthias Vallentin. After you finish with essentials of mathematics, this book will help you connect various theorem and algorithm quickly with their formulae. It's difficult to derive equations instantly, this book will help you to quickly navigate to your desired problem and solve. Available: Free Download

End Notes

The books listed here are selected on the basis of their reviews and depth of topics covered. This is not an exhaustive list of books. But, we found it's almost too easy to get confused while deciding 'from where to begin?' In such situations, it is advisable to start with this list.

It has been found that people tend of neglect these topics in pursuit of quick success. But, that's not the right way. Hence, if you aim for a long term success in data science, make sure you learn to create stories out of maths and statistics.

Source: Analytics Vidhya