Foreword

The 21st century will probably be the century of the data revolution. Our numerical world is creating masses of data every day and the volume of generated data is increasing more and more (the number of produced numerical data is doubling every two years according to the most recent estimates). In such a context, data science is nowadays an unavoidable field for anyone interested in exploiting data. People may be interested in either understanding a phenomenon or in predicting the future behavior of this phenomenon.

To this end, it is important to have significant knowledge of both the rationale (the theory) behind data science techniques and their practical use on real-world data. Indeed, data science is a mix of data, statistical/machine learning methods and software. Software is actually the link between data and data science techniques. It allows the practitioner to load the data and apply techniques on it for analysis. It is therefore important to master at least one of the data science languages.

The choice of the software language(s) mainly depends on your background and the expected level of analysis. R and Python are probably the two most popular languages for data science. On the one hand, R has been made by statisticians. . . mostly for statisticians! It is, however, an excellent tool for data science since the most recent statistical learning techniques are provided on the R platform (named CRAN). Using R is probably the best way to be directly connected to current research in statistics and data science through the packages provided by researchers. Python is, on the other hand, an actual computer science language (with all appropriate formal aspects) for which some advanced libraries for data science exist. In this context, the Julia language has the great advantage to permit users to interact with both R and Python (but also C, Fortran, etc.), within a software language designed for efficient and parallel numerical computing while keeping a high level of human readability.

Professor Paul McNicholas and Peter Tait propose in this book to learn both fundamental aspects of data science: theory and application. First, the book will provide you with the significant elements to understand the mathematical aspects behind the most used data science techniques. The book will also allow you to discover advanced recent techniques, such as probabilistic principal components analysis (PPCA), mixtures of PPCAs, and gradient boosting. In addition, the book will ask you to dive into the Julia language such that you directly apply the learned techniques on concrete examples. This is, in my opinion, the most efficient way to learn such an applied science. In addition, the focus made by this book on the Julia language is a great choice because of the numerous qualities of this language regarding data science practice. These include ease of learning for people familiar with R or Python, nice syntax, easy code debugging, the speed of the compiled language, and code reuse.

Both authors have extensive experience in data science. Professor Paul McNicholas is Canada Research Chair in Computational Statistics at McMaster University and Director of the MacDATA Institute of the same university. In his already prolific career, McNicholas has made important contributions to statistical learning. More precisely, his research is mainly focused on model-based learning with high-dimensional and skew-distributed data. He is also a researcher deeply involved in the spreading of research products through his numerous contributions to the R software with packages. Peter Tait is currently a Ph.D. student but, before returning to academia, he had a professional life dedicated to data science in industry. His strong knowledge of the needs of industry regarding data science problems was really an asset for the book.

This book is a great way to both start learning data science through the promising Julia language and to become an efficient data scientist.

Professor Charles Bouveyron
Professor of Statistics
INRIA Chair in Data Science
Université Côte d'Azur
Nice, France