

Statistical Learning from a Regression Perspective

Richard A. Berk

Springer, 2008, xviii + 360 pages, £ 55.99 / € 69.95 / US\$ 84.95, hardcover

ISBN: 978-0-387-77500-5

Table of contents

- | | |
|---|---|
| 1. Statistical learning as a regression problem | 5. Random forests |
| 2. Regression splines and regression smoothers | 6. Boosting |
| 3. Classification and regression trees (CART) | 7. Support vector machines |
| 4. Bagging | 8. Broader implications and a bit of craft lore |

Readership: Graduate students and researchers in the social and life sciences, and biomedical and ecological sciences.

As the title indicates, the focus of this book is on statistical learning procedures that can be understood within a regression framework. I believe that the practical utility of statistical learning over more traditional non- and parametric regression approaches has yet to be truly demonstrated but the procedures presented in this text do show considerable potential.

Statistical learning is currently undergoing rapid growth and change, and one must evaluate all assertions about performance with considerable care. The text is suitable for private study but it is imperative that the limitations of one's data are understood in order not to apply inappropriate statistical learning procedures.

The mathematical prerequisites for using this book are minimal. A number of the methods are presented in the context of the freeware software R – a programming language designed for statistical computing and graphics. There are exercises at the end of each chapter which are structured around R. Some familiarity with using a computer is necessary in order to gain the most benefit from the text, and some previous experience of using a statistical software package would be advantageous.

C.M. O'Brien: carl.obrien@cefas.co.uk

Centre for Environment, Fisheries & Aquaculture Science
Pakefield Road, Lowestoft, Suffolk NR33 0HT, UK

Meta-analysis of Binary Data Using Profile Likelihood

Dankmar Böhning, Ronny Kuhnert, Sasivimol Rattanasiri

Chapman & Hall/CRC, 2008, xv + 190 pages, £ 44.99 / US\$ 79.95, hardcover

ISBN: 978-0-387-77500-5

Table of contents

- | | |
|---|---|
| 1. Introduction | 7. Working with CAMAP |
| 2. The basic model | 8. Estimation of odds ratio using the profile likelihood |
| 3. Modeling unobserved heterogeneity | 9. Quantification of heterogeneity in a MAIPD |
| 4. Modeling covariate information | 10. Scrapie in Europe: a multicountry surveillance study as a MAIPD |
| 5. Alternative approaches | |
| 6. Incorporating covariate information and unobserved heterogeneity | |

Readership: Students of applied statistics and biometrics.

In some areas of scientific investigation, meta-analysis has become both an essential method and a standard tool of analysis. This book presents a unifying approach to modelling treatment effect in a meta-analysis of clinical trials with binary outcome.

The profile log-likelihoods that are used in this book are non-standard – neither a Poisson log-likelihood nor any of the log-likelihoods available in the generalized linear model family. As a consequence, a software tool CAMAP (Computer Assisted Analysis of Meta-Analysis using the Profile Likelihood Model) has been developed for estimating the relative risk based upon the profile log-likelihood models. The authors describe how to work with CAMAP and the authors state that software can be downloaded with no costs from www.reading.ac.uk/sns05dab/software.html, although at the time of writing this site was inaccessible. However, the site www.personal.reading.ac.uk/~sns05dab/Software.html is accessible and provides the desired access.

The text contains many real-world examples which add to the usefulness of the book. Included amongst these are trials on cancer treatments, diseased cows, beta-blockers and hypertension, to mention but a selection. The balance between statistical theory and practical applications with CAMAP make the text suitable for private study and research.

C.M. O'Brien: carl.obrien@cefas.co.uk

Centre for Environment, Fisheries & Aquaculture Science
Pakefield Road, Lowestoft, Suffolk NR33 0HT, UK

Bioconductor Case Studies

Florian Hahne, Wolfgang Huber, Robert Gentleman, Seth Falcon
Springer, 2008, xii + 284 pages, £ 41.99 / € 52.95 / US\$ 64.95, softcover
ISBN: 978-0-387-77239-4

Table of contents

- | | |
|--|--|
| 1. The ALL dataset | 9. Supervised machine learning |
| 2. R and bioconductor introduction | 10. Unsupervised machine learning |
| 3. Processing affymetrix expression data | 11. Using graphs for interactome data |
| 4. Two-color arrays | 12. Graph layout |
| 5. Fold-changes, log-ratios, background correction, shrinkage estimation, and variance stabilization | 13. Gene set enrichment analysis |
| 6. Easy differential expression | 14. Hypergeometric testing used for gene set enrichment analysis |
| 7. Differential expression | 15. Solutions for exercises |
| 8. Annotation and metadata | |

Readership: Students and applied researchers of statistics, computer science, bioinformatics and molecular biology.

This book comprises a collection of co-authored individual chapters resulting from the annual Bioconductor courses held in Bressanone, Italy. Each chapter focuses on a case that uses real data to illustrate specific problems in genomics and molecular biology.

The readership for this book will be specialised but the text deserves to be read more widely within the statistics and computer science communities as there is much to interest