Book Reviews

Editor: Chaitra H. Nagaraja

Surrogates: Gaussian Process Modeling, Design, and Optimization for the Applied Sciences
Robert B. Gramacy
Chapman & Hall/CRC, 2020, 559 pages, £71.99, hardcover
ISBN: 978-0-367-41542-6

Readership: Graduate students or professional students.

In his book *Surrogates: Gaussian Process Modeling, Design, and Optimization for the Applied Sciences*, Gramacy focuses on the use of Gaussian processes to problems of interest to researchers considering design of experiments in their work. He proceeds in somewhat reverse fashion: the problems in the design of experiments are presented earlier in the book, followed by the Gaussian process framework, estimation and inference approaches and attendant computational issues.

The term 'surrogate' is also referred to as 'emulator' in the statistical literature (e.g. Sacks *et al.*, 1989; Kaufman *et al.*, 2011), and, ideally, the surrogate has the following properties:

1. the predictive distribution of the surrogate is what will be used for inference;
2. the models for the surrogate should have good interpolation properties for the original data-generating process;
3. the surrogate can be used in the same way the model for the original data-generating process could have been used, with appropriate uncertainty quantification; and
4. the model for the surrogate is easier to work with than that for the original data-generating process.

The book is a fusion of methods for experimental design and methodology for Gaussian processes which traverses the fields of machine learning, spatial statistics and computational algorithms. These represent large areas of statistics that the author has to cover and summarise succinctly. He is successful in doing so, conveying essential ideas in the main body of the text and providing references for interested readers wanting more specialised knowledge in the subfields that use Gaussian processes.

The book takes an example-based approach to a variety of problems which are interlaced with R code. For each chapter, there is a set of exercises to test learners on the material. While not explicitly mentioned, readers are assumed to have a familiarity with the R language, as the code in the text is all R based. The author has a very conservational writing style with a pretty dry sense of humour that, in conjunction with the examples, makes the material very accessible to students learning the material for the first time. The other aspect I liked about the material in Chapters 1 through 4 was the use of visualisation rather than statistical output in order to illustrate statistical ideas.

While Chapters 1 through 4 deal with topics in experimental design, Chapters 5 through 10 deal with Gaussian processes, which the author has researched extensively during his career. In addition to the articles Gramacy has published, he has developed a suite of R packages to implement the methods in these chapters. This provides an immeasurable boost to the power and impact of the book. Chapters 5 through 10 cover the main substance in terms of fitting the Gaussian process models from both a conceptual and a practical point of view. The author provides a nice historical description on the use of Gaussian processes in statistics and machine learning. While he adopts a Bayesian approach to inference, it is borne out of pragmatism. As such, Chapter 5 focuses on inference for posterior distributions using Gaussian processes. It also goes in-depth on issues of prior selection, computational implementations and covariance function selection, which are key issues in modelling using Gaussian processes. Chapter 6 deals with designs based on fitted Gaussian processes, including ideas from active learning, in which design points are selected based on model predictions from the surrogate. These ideas are then applied to numerical optimisation problems in Chapter 7. In Chapter 8, the author describes the application of Gaussian process models for calibration experiments, an extremely active area of current statistical research. This has been an important modelling technique for a variety of problems, such as climate change models (Chang *et al.*, 2016), materials science (Zhang *et al.*, 2020) and other physical systems (Li *et al.*, 2020). While there is a natural progression in Chapters 5 through 7 in terms of the topics being developed and then applied, Chapters 9 and 10 pivot to issues of computational scalability and flexibility. In Chapter 9, a variety of computational approaches to handle large-scale datasets are described. These come from a variety of fields, including applied mathematics (compactly supported kernels), machine learning and spatial statistics (predictive processes) and the tree models that the author and other researchers have developed and expanded over the last 15 years. Chapter 10 focuses on developing Gaussian process models that model nonstationarity but that can scale with big datasets. Due to the active research on the topics in Chapters 9 and 10, I envision that this material would be updated in future versions of the book.

In conclusion, *Surrogates: Gaussian Process Modeling, Design, and Optimization for the Applied Sciences* is a book that is a fusion of response surface methodology and associated problems with Gaussian process modelling. Gramacy covers a lot of ground while being very attentive to the various fields of machine learning and statistics that have considered Gaussian processes. He has synthesised the knowledge about these topics in a very interesting and fresh manner. The book is a great introduction to Gaussian processes and their use on large-scale datasets, along with their application to various problems in design of experiments. The R code provided will allow users of the book to be able to implement these methods quickly in practice. I look forward to future editions of the book.

References

Chang, W., Haran, M., Applegate, P. & Pollard, D. (2016). Calibrating an ice sheet model using high-dimensional binary spatial data. *J. Amer. Stat. Assoc.*, **111**, 57–72.

Kaufman, C.G., Bingham, D., Habib, S., Heitmann, K. & Frieman, J.A. (2011). Efficient emulators of computer experiments using compactly supported correlation functions, with an application to cosmology. *Ann. Applied Stat.*, **5**, 2470–2492.

Li, K., Wen, J., Liu, Y., Liu, H., Wang, S. & Tu, J. (2020). Application of entransy theory on structure optimization of serrated fin in plate-fin heat exchanger. *Appl. Therm. Eng.*, **173**, 114809.

Sacks, J., Welch, W.J., Mitchell, T.J. & Wynn, H.P. (1989). Design and analysis of computer experiments. *Statist. Sci.*, **4**, 409–423.

Zhang, Y., Apley, D.W. & Chen, W. (2020). Bayesian optimization for materials design with mixed quantitative and qualitative variables. *Sci. Rep.*, **10**, 4924.

Debashis Ghosh: *Debashis.ghosh@cuanschutz.edu*
Department of Biostatistics and Informatics
Colorado School of Public Health
13001 E. 17th Place, University of Colorado Anschutz Medical Campus, Aurora, CO 80045,
USA

Artificial Intelligence for Drug Development, Precision Medicine, and Healthcare

Mark Chang

Chapman & Hall/CRC, 2020, 368 pages, £35.99, ebook

ISBN: 978-0-367-36292-8

Readership: Practitioners of statistical science in healthcare applications.

This book would be a neat addition to the practitioner's reference library of statistical methodologies for healthcare data analysis based on artificial intelligence (AI) and machine learning. The technical narrative is written in recipes; these are sandwiched between a high-level introduction (comprising modern AI and machine learning, classical statistics, and the similarity principle) and an epilogue espousing the author's perspectives about future progress in modern AI. The statistical recipes include example R programming code for the following methodologies: artificial and deep learning neural networks, kernel methods, decision tree and ensemble methods, Bayesian learning, unsupervised learning, reinforcement learning, and swarm and evolutionary intelligence. The penultimate chapter concisely accounts for modern applications of AI in medical science and drug development with a wide scope, which is a unique characteristic of this book. The experienced practitioner would appreciate the organization of technical content and the contextual narrative; to make most use of this content, they should be familiar with the underlying statistical concepts in machine learning. As such, this book would not be practical for classroom instruction. It is a compactly written book that could serve as a handy reference guide (i.e., cookbook) for the practitioner who would need to quickly review a new methodology and understand the bigger picture as it would relate to applications in healthcare.

Frank Yoon: *fyoon@us.ibm.com*
Government Health and Human Services, IBM Watson Health 75 Binney Street,
Cambridge, 02142, MA, USA

Exploratory Subgroup Analysis in Clinical Research

Gerd Rosenkranz

Wiley, 2020, 248 pages, $90, hardcover

ISBN: 978-1-119-53697-0

Readership: graduate students, clinicians and statistics researchers.

This is a useful book, which covers different methods to perform subgroup analysis. It is very well written and easily understood by a broad audience without prior knowledge of subgroup

analysis. This book focuses on classical regression models like (generalized) linear models and the proportional hazards model for survival data and emphasizes applications with relatively few biomarkers or covariates. Clinicians and statisticians from the area of clinical development and regulation would benefit most, although the topic of subgroup analysis has a much wider scope.

The book is very well structured. First, it describes the history of subgroup analyses and introduces historical subgroup analyses which are each remarkable for a special reason. Next, it defines subgroups and presents examples of the good, the bad, and the ugly subgroup analyses from the past, followed by different methods to analyze data from subgroups, including hierarchical models, methods to obtain estimates of expected individual treatment effects and prediction models (Chapters 4–8). Each method is illustrated with case studies. This book also carefully addresses selection bias by the bootstrap method and the risk of selecting too many false positive results in subgroup analyses.

Overall, it is an exciting book, and it would provide an excellent basis for subgroup analyses.

Lili Zhao: *zhaolili@umich.edu*
Department of Biostatistics, University of Michigan, 1415 Washington Heights,
Ann Arbor, MI, 48109, USA

Statistical Machine Learning: A Unified Framework
Richard M. Golden
Chapman and Hall/CRC, 2020, 524 pages, £71.99, hardcover
ISBN: 978-1-138-48469-6

Readership: This textbook is not only suitable for graduate students but also useful for professionals, scientists and others in the areas of statistics, applied mathematics, computer science and/or electrical engineering.

Several statistical learning textbooks have been written to provide comprehensive descriptions of algorithms, use of software tools or support and evaluation of machine learning architectures. From a varying perspective, this textbook presents a statistical machine learning framework based on the assumption that machine learning algorithms learn a best-approximating probability distribution for representing the true data generating process.

This textbook has four parts.

Part 1 introduces the concept of machine learning algorithms by intuitive examples.

Chapter 1 illustrates that many supervised, unsupervised and reinforcement learning algorithms may be viewed as empirical risk function optimisation algorithms. Chapter 2 uses set theory to represent concepts and logical formulas. It also uses metric spaces to represent concept similarity. Chapter 3 discusses how to model both discrete-time and continuous-time learning

environments for machine learning algorithms, defines the computational goal and interprets the algorithms as rational decision-making machines.

Part 2 focuses on deterministic learning machines.

Chapters 4 and 5 cover those essential results in linear algebra and matrix calculus. Chapter 6 provides ample conditions for characterising the asymptotic behaviour of discrete-time and continuous-time time-invariant dynamical systems. Chapter 7 provides sufficient conditions for ensuring that a large class of deterministic batch learning algorithms converge to the set of critical points of the objective function for learning.

Part 3 is concerned with stochastic inference and stochastic learning machines.

Chapters 8–10 discuss random vectors, stochastic sequences and probability models needed in machine learning. Chapter 11 develops the asymptotic convergence theory for Monte Carlo Markov chains for the special case where the Markov chain is defined on a finite state space. Chapter 12 provides relevant asymptotic convergence analyses of adaptive learning algorithms for both passive and reactive learning environments.

Part 4 is devoted to the problem of characterising the generalisation performance of a machine learning algorithm.

Chapter 13 first discusses the risk functions and then maximum likelihood and maximum a posterior estimation methods. Chapters 14–16 demonstrate how both bootstrap simulation methods and asymptotic formulas can be used to characterise the generalisation performance of the class of machine learning algorithms considered.

This textbook includes a self-contained introduction to mathematical and technical details. The introductions to real analysis (Chapter 2), linear algebra (Chapter 4), matrix calculus (Chapter 5), measure theory (Chapter 8) and stochastic sequences (Chapter 9) cover those essential materials and add value to the required mathematical prerequisites. For example, Chapter 9 includes the definitions of convergence in mean square, convergence in probability and convergence in distribution, and their relationships as well as the Uniform Law of Large Numbers and Weak Law of Large Numbers. A special mention should be given to Chapter 5 which provides a useful coverage of advanced matrix calculus results to support machine learning applications. The definition of a matrix derivative is discussed, with a specific focus on its applications within machine learning and deep learning.

This textbook also includes explicit examples and insightful discussions of machine learning applications. For example, Chapter 1 discusses image processing, neural networks and clustering examples when feature vectors, supervised learning and unsupervised learning are introduced. Chapter 12 examines pertinent applications of the Gibbs sampler, the

Metropolis–Hastings algorithm, minibatch stochastic gradient descent, stochastic approximation expectation maximisation and policy gradient reinforcement adaptive learning. Chapter 16 looks at model selection and evaluation from a machine learning point of view with a Bayesian hypothesis testing example.

Some advice and encouragement and suggestions for reading the book are given in the preface. Example course syllabi, sample exams, supplementary instructional materials, software problems and book revisions are promised to be made available on the author's website: www.statisticalmachinelearning.com.

Shuangzhe Liu: *shuangzhe.liu@canberra.edu.au*

Faculty of Science and Technology, University of Canberra,
Canberra, ACT 2602, Australia