Chapter 11 includes more case studies in other areas, ranging from manufacturing to marketing research. Chapter 12 concludes the book with some commentary about the scientific contributions of MTS.

The Taguchi method for design of experiment has generated considerable controversy in the statistical community over the past few decades. The MTS/MTGS method seems to lead another source of discussions on the methodology it advocates (Montgomery 2003). As pointed out by Woodall et al. (2003), the MTS/MTGS methods are considered ad hoc in the sense that they have not been developed using any underlying statistical theory. Because the "normal" and "abnormal" groups form the basis of the theory, some sampling restrictions are fundamental to the applications. First, it is essential that the "normal" sample be uniform, unbiased, and/or complete so that a reliable measurement scale is obtained. Second, the selection of "abnormal" samples is crucial to the success of dimensionality reduction when OAs are used. For example, if each abnormal item is really unique in the medical example, then it is unclear how the statistical distance MD can be guaranteed to give a consistent diagnosis measure of severity on a continuous scale when the larger-the-better type S/N ratio is used.

Multivariate diagnosis is not new to *Technometrics* readers and is now becoming increasingly more popular in statistical analysis and data mining for knowledge discovery. As a promising alternative that assumes no underlying data model, *The Mahalanobis–Taguchi Strategy* does not provide sufficient evidence of gains achieved by using the proposed method over existing tools. Readers may be very interested in a detailed comparison with other diagnostic tools, such as logistic regression and tree-based methods.

Overall, although the idea of MTS/MTGS is intriguing, this book would be more valuable had it been written in a rigorous fashion as a technical reference. There is some lack of precision even in several mathematical notations. Perhaps a follow-up with additional theoretical justification and careful case studies would answer some of the lingering questions.

Wei JIANG
AT&T Laboratories

REFERENCES

Krzanowski, W. J. (1977), "The Performance of Fisher's Linear Discriminant Function Under Non-Optimal Conditions," *Technometrics*, 19, 191–200.

Montgomery, D. (2003), "Book Review: *The Mahalanobis–Taguchi Strategy*," *Journal of Quality Technology*, 35, 229–231.

Taguchi, G., Chowdhury, S., and Wu, Y. (2001), *The Mahalanobis–Taguchi System*, New York: McGraw-Hill.

Woodall, W. H., Koudelik, R., Tsui, K.-L., Kim, S. B., Stoumbos, Z. G., and Carvounis, C. P. (2003), "A Review and Analysis of the Mahalanobis–Taguchi System" (with discussion), *Technometrics, 45, 1–30.*

The Elements of Statistical Learning, by Trevor HASTIE, Robert TIBSHIRANI, and Jerome FRIEDMAN, New York: Springer-Verlag, 2001, ISBN 0-387-95284-5, xvi + 533 pp., $74.95.

Occasionally a book comes along that I decide I need to read, so I make myself the reviewer. My choices are not always wise. I need to be able to read a book quickly. This was definitely not an easy book to read. But I felt that the topics that it covers were important for me to know, so I persevered. Subtitled "Data Mining, Inference and Predictions," this book is the first attempt by statisticians to gather the tools of the modern methods for learning from data into a single textbook. A significant portion of the theoretical development for this methodology has been done by these authors, (see, e.g., Hastie and Tibshirani 1991; De Veaux 1992; Friedman and Stuetzle 1981; Friedman 1991).

Unfortunately for the practitioner, or the casual reader, the high level of mathematical scholarship exemplified by these authors is very evident within this book. Although the book was intended "for researchers and students in a broad variety of fields" (p. 7) and presumes only that "the reader will have had at least one elementary course in statistics," that one course must have given the reader a very solid background, particularly in regression modeling and decision theory. The reader will be most appreciative of the book after the standard 1-year sequences in both mathematical statistics and applied statistics, plus some experience with data analysis for large amounts of data.

Chapter 1, "Introduction," gives some of the examples that the authors use as the motivation for reading the book. Chapter 2, "Overview of Supervised Learning," is the background for all of the rest of the book except the final chapter. Unfortunately, it contains no examples, only mathematical statements about different prediction methods, decision theory, various local estimation methods, function approximation, and restricted estimators. For statisticians, this is important theoretical background, but it does not facilitate practical learning. The authors describe their approach as follows: "Treating supervised learning as a problem in function approximation encourages geometrical concepts of Euclidean spaces and mathematical concepts of probabilistic inference to be applied to the problem" (p. 30). That statement pretty much exemplifies the tone of the entire book as well.

Chapters 3 and 4, "Linear Methods for Regression" and "Linear Methods for Classification," are also necessary background for the rest of the book. Considerable effort is expended in explaining the mathematics, geometry, and inferences for estimation in least squares regression. This would be review for well-educated statisticians, but it is certainly also couched in their particular language. Although there is an example that appears a couple of times, the theoretical constructs are certainly at least equally as important in these chapters as the applications for the methods. The writing is uniformly excellent. Explanations of different shrinkage methods, for example, are both mathematically and verbally explicit. Extensive use of graphics makes this particular differentiation visual as well. The computational considerations at the ends of these chapters deal with numerical methods, not software. The heavily theoretical approach is manageable for the relatively familiar regression methodology, but quickly becomes overwhelming for the less-familiar classification methods. An example on heart disease helps out, but examples are not this book's primary focus. A comparison of logistic regression and linear discriminant analysis is informative. Trying to follow the subsequent discussion on hyperplane classifiers is challenging.

The first four chapters and Chapter 7, "Model Assessment and Selection," are the chapters considered mandatory reading, because they "cover central concepts that pertain to all learning methods" (p. 7). The chapters in between cover "Basic Expansions and Regularization" and "Kernel Methods." The first of these two chapters deals with splines and wavelets, two methods that are heavy on the mathematics. The linear regression and discriminant analysis examples provided helpful interludes from the methodology, but this is more material that is very challenging to read. Examples here provide mostly visual results. Material on logistic regression and multidimensional splines serves as an introduction to subsequent material on generalized additive modeling and the multivariate adaptive regression splines (MARS) procedure. Several sections, including "Computational Considerations," are considered "technically difficult" (p. 8).

Chapter 6, "Kernel Methods," is another difficult background chapter. It is heavily focused on the computational statistics of local regression. This is probably very useful for parameterizing functions in S-PLUS, the principal computational tool, which the authors use to actually implement these procedures with data. Kernel density estimation is also discussed. Model selection in Chapter 7 provides the tools for choosing a learning model and determining the quality of fit. This process focuses on the use of test samples. Appropriate loss functions are needed to quantify the errors, and the chapter emphasizes methods for estimating a curve for test error versus model complexity. The authors note the idealized situation of dividing the data into training, validation, and test data sets (e.g., their recommended 50%–25%–25% split), but they focus on methods with enough data for only one division. The tools that are used are either analytical, such as the Akaike information criterion, or sample reuse, such as bootstrapping. This is all mostly mathematics, but there is enough text so that I could comprehend what was going on.

Chapter 8, "Model Inference and Averaging," actually is another background chapter, which completes the first half of the book. This chapter focuses on maximum likelihood estimation and Bayesian inference and their relationship to the bootstrap. It starts with a smoothing example and uses both nonparametric and parametric bootstraps. Two middle sections of the chapter compare and contrast the EM algorithm and the Markov chain Monte Carlo approach to posterior sampling, represented by the well-known Gibbs sampler. Esoterica prevails in the discussions of bagging, stacking, and bumping that conclude the chapter. Somewhere John Tukey must be grinning.

All of the foregoing was the background! The first actual methods for supervised learning are the generalized additive models (GAMs) in Chapter 9 and their assorted friends: tree-based methods, MARS, and hierarchical mixtures of experts. GAMs are illustrated using additive logistic regression. Computational algorithms are succinctly stated in text boxes. Here readers are directed

to specialized software. For tree-based methods, the Chapter describes CART and C5.0, plus a variation known as patient rule induction methods or bump hunting. A common example is used. The situation with missing data and the computational effort necessary for each method are also considered.

Chapter 10, "Boosting and Additive Trees," describes boosting as "one of the most powerful learning ideas introduced in the last ten years" (p. 299). The book develops boosting methods for classifiers and extends them to regression applications. The method called "Ada Boost" is described and used for additive models. Boosting is "a way of fitting an additive expansion in a set of elementary basis functions" (p. 310). Though this material is very advanced, the chapter discusses in a couple of sections the creating of "off-the-shelf" (p. 312) data-mining procedures using predictive learning methods. The authors note that "requirements of speed, interpretability and the messy nature of the data sharply limit the usefulness of most learning procedures as off-the-shelf methods for data mining" (p. 313). Decision trees are determined to be the best tool available. Boosting methods are recommended to improve accuracy via a multiple adaptive regression tree (MART). MART is illustrated for two large public-domain datasets.

The statistician's approach to neural networks (NNs) is the subject of Chapter 11. Projection pursuit regression (PPR) provides the starting point for the presentation. The single-layer perceptron is the NN that has been selected for discussion. The authors note that NNs "are just nonlinear statistical models, much like the PPR" (p. 350). It was gratifying to see what easy work the authors could make of the entire process of configuring an NN with the background of the book's first 350 pages, although the authors still note that "there is quite an art in training neural networks" (p. 355). Several pages of guidance, along with two extensive examples, are provided.

Chapter 12 offers generalizations of the use of linear decision boundaries for classification. Techniques discussed include support vector machines (SVMs) and flexible discriminant analysis, the latter also including penalized discriminant analysis and mixture discriminant analysis. Applications of SVMs include regression analysis. This material is very complex, but some nice graphics and basic examples aid understanding. Readers are directed to S-PLUS programs. Chapter 13 continues along a similar vein with other methods for classification and pattern recognition. These model-free methods are touted as "black box prediction engines" (p. 411). These include prototype methods, such as K-means clustering or k-nearest-neighbor classifiers, and adaptive methods.

A chapter on unsupervised learning concludes the book. Here the link to all of the supervised methods that precede this chapter is very enlightening. The chapter discusses various association rules, including market basket analysis, cluster analysis, self-organizing maps, principal components, independent component analysis (ICA), and multidimensional scaling. The 40 pages on cluster analysis include discussion of many algorithms, including combinatorics, K-means, vector quantization, K-medoids, hierarchical clustering, agglomerative clustering, and divisive clustering. Readers interested in ICA should investigate the recent book by Roberts and Everson (2001), reviewed by Rayens (2003).

The Elements of Statistical Learning is a vast and complex book. Generally, it concentrates on explaining why and how the methods work, rather than how to use them. Examples and especially the visualizations are principal features, but little guidance is available to reader who would want to reproduce these results. As a resource for the methods of statistical learning, however, it will probably be a long time before there is a competitor to this book.

Eric R. Ziegel
BP

REFERENCES

De Veaux, R. (1992), Review of *Generalized Additive Models*, by T. Hastie and R. Tibshirani, *Technometrics*, 34, 225–226.

Friedman, J. (1991), "Multivariate Adaptive Regression Splines" *The Annals of Statistics*, 19, 1–141.

Friedman, J., and Stuetzle, W. (1981), "Projection Pursuit Regression," *Journal of the American Statistical Association*, 76, 817–823.

Hastie, T., and Tibshirani, R. (1991), *Generalized Additive Models*, New York: Chapman & Hall.

Rayens, W. (2003), Review of *Independent Component Analysis*, by S. Roberts and R. Everson, *Technometrics*, 45, 107–108.

Roberts, S., and Everson, R. (eds.) (2001), *Independent Component Analysis*, Cambridge, U.K.: Cambridge University Press.

Elements of Computational Statistics, by James E. Gentle, New York: Springer-Verlag, 2002, ISBN 0-387-95489-9, xviii + 420 pp., $79.95.

This book is an audacious undertaking by the author—an effort to present all of the major statistical methods that require a large degree of computational intensity. Gentle states that his intent is to "describe these methods in a general manner and to emphasize the commonalities among them" (p. viii). Many topics are areas of current research, and to present this clearly within the boundaries of a reasonably sized text is a challenge. Because of this restriction, the chapters are concise, and proofs and many details are not given. However, the author does include an extensive bibliography (as well as "Further Reading" sections at the end of many chapters) that allows the reader to research most topics. For the probable purchasers of this text, I feel that Gentle has succeeded in presenting a broad overview of the major areas of modern computational statistics.

I should point out that this book is not about statistical computing, which was the subject matter of an earlier text by Kennedy and Gentle (1980). Gentle makes the distinction between *computational statistics* and *statistical computing* in the first paragraph of the Preface. He views computation statistics as "the class of statistical methods characterized by computational intensity" (p. vii), whereas statistical computing involves computational methods applied to statistics, for example, various areas of numerical analysis. Gentle gives references for those more interested in the latter, but he mistakenly overlooks the excellent book by Thisted (1988), a text with a similar title to this one.

The book comprises 11 chapters of varying length (e.g., Chap. 8 is only six pages long). Each chapter ends with a small set of challenging exercises, and the book is intended as a textbook for a graduate-level course in computational statistics (or even at the advanced undergraduate level). However, given that the chapters are brief and that many details are omitted, it may be the case that the text could be quite difficult to lecture from. Gentle claims in the Preface that when he teaches the material, he includes more extensive examples for class lectures, and these are available on the text's web page. (To date, these have not been made available.)

The book is divided into two parts. Part I, "Methods of Computational Statistics" (Chaps. 1–7), includes preliminary material necessary for statistical inference and Monte Carlo (MC) techniques. I am pleased how Gentle explicitly presents the difference between the random variable and a *realization* of it. He carries this throughout the text when he distinguishes *estimators* from *estimates* (of parameters). Part I also includes such topics as the bootstrap and graphical methods. I Part II, "Exploring Data Density and Structure" (Chaps. 8–11), covers density estimation along with clustering and classification techniques used in multivariate data analysis (MDA). The remainder of the text includes four appendixes that contain a discussion on MC experiments and software (devoted primarily to S-PLUS and R), a section on notations and definitions, and hints and solutions to some problems. In addition, Gentle maintains a web page of errata for this book (which identified most of the errors that I discovered while reviewing this text).

Chapters 1 and 2 are the introductory chapters that present topics of inference and MC methods and explain most of the notation used in the book. In Chapter 1, much ground is covered in the areas of mathematical statistics, linear models, and advanced methods of statistical analysis. Gentle stresses the development of the empirical cumulative distribution function (ECDF) and explains the importance of its role in estimation as well as methods in computational statistics (e.g., the bootstrap). In addition, many methods of estimation (e.g., least squares and those based on the likelihood) are discussed and are portrayed as *optimization* procedures in inference. It is here that Gentle introduces the EM method and includes the nice example using missing data from Flury and Zoppè (2000). Chapter 2 introduces MC methods that at their core rely on simulating from distribution functions. Logically, Gentle begins with a description of the generation of pseudorandom numbers using standard techniques and Gibbs sampling. Other MC techniques, such as MC estimates of definite integrals and Markov chain models, are mentioned only briefly. This chapter moves very quickly, so a reader unfamiliar with many of these methods may need additional resources.

Chapters 3 and 4 are two concise chapters that introduce resampling methods used in inference. Discussed herein are (primarily) the jackknife and the bootstrap. In Chapter 3, "Randomization and Data Partitioning," Gentle does a