
BOOK REVIEWS¹

The Elements of Statistical Learning: Data Mining, Inference and Prediction. Trevor Hastie, Robert Tibshirani, and Jerome Friedman. New York: Springer-Verlag, 2001, 533 pages.

Generalizability Theory. Robert L. Brennan. New York: Springer-Verlag, 2001, 538 pages.

Reviewed by George A. Marcoulides
California State University, Fullerton

THE ELEMENTS OF STATISTICAL LEARNING: DATA MINING, INFERENCE, AND PREDICTION

There is much current research in the statistics community on developing and using algorithms for discovering knowledge and structure from vast amounts of data. Some researchers call this "learning from data" (e.g., Bradshaw & Boose, 1990). Others simply refer to it as "data mining" (e.g., Padmanabhan & Tuzhilin, 2002). Data mining is at the interface between statistics and information (computer) systems. It is an emerging field that has attracted much attention in a very short period of time. Data mining is applicable across a variety of disciplines and provides methodologies for such diverse problems as fraud detection, customer satisfaction, risk analysis, bankruptcy prediction, and credit scoring, to name a few.

This book provides an excellent introduction to learning from data. It is written at a level that should appeal not only to statisticians but also to researchers and practitioners from a wide variety of fields. Although some mathematical details are provided, the emphasis tends to be more on the conceptual aspects of each method rather than their theoretical properties. There are also some very good example cases throughout the various chapters. The book concentrates mainly on methods that are considered *supervised* learning, but does offer one chapter on *unsupervised* learning. In general, supervised learning attempts to predict the value of an outcome variable using any number of input variables. Based on this definition, various types of discriminant and regression analyses approaches would be

Requests for reprints should be sent to George Marcoulides, Department of Information Systems and Decision Science, Langsdorf Hall 540, California State University, Fullerton, CA 92634-9480.

¹Three books are reviewed in this section. Due to space limitations, full comprehensive reviews were not possible.

categorized as supervised learning. In contrast, with unsupervised learning the focus is on discovering and describing associations and patterns among variables. As such, cluster analysis, factor analysis, and various latent variable models would be categorized as unsupervised learning. Although the topics covered in this book do not exhaust the range of data mining techniques, the ones selected do represent some of the most commonly used techniques. In sum, there is much to recommend about this book.

REFERENCES

- Bradshaw, J. M., & Boose, J. H. (1990). Decision analysis techniques for knowledge acquisition: Combining information and preferences using Aquinas and Axotl. *International Journal of Man-Machine Studies*, 32, 121-186.
- Padmanabhan, B., & Tuzhilin, A. (2002). Knowledge refinement based on the discovery of unexpected patterns in data mining. *Decision Support Systems*, 33, 309-321.