readability could have been improved considerably if matrices and vectors had been set in boldface. Nevertheless, the overall evaluation is clearly positive. The book is indeed very comprehensive and should be a valuable addition to the researcher's collection.

Bernhard Flury
Indiana University

Algorithms for Clustering Data
by Anil K. Jain and Richard C. Dubes
Prentice Hall Advanced Reference Series in Computer Science
Prentice Hall, Englewood Cliffs, NJ, 1988

Quite a few books on cluster analysis are currently available, but I think that this one is indeed a valuable addition. It has been written for computer scientists or engineers rather than for statisticians, and requires relatively moderate mathematical prerequisites. The book is generally not difficult to read, gives good explanations of the methods, and provides nice examples.

After a brief introduction (Chapter 1) to the purpose of cluster analysis, Chapter 2 discusses data types and scales, proximity indices, normalization and projection of data, intrinsic (or topological) dimensionality of patterns in high-dimensional space, and multidimensional scaling. Chapter 3, the main chapter of the book, treats the various clustering methods and algorithms. These are divided into hierarchical methods and partitional methods, the latter ones including mixture decomposition, density estimation, and fuzzy clustering. Chapter 4, on 'cluster validity', treats the important (and often ignored) problem of evaluating the results of a cluster analysis in a quantitative fashion, using (at least in part) statistical reasoning. Chapter 5 presents applications of clustering techniques to image processing. Several appendices review briefly such diverse topics as pattern recognition, selected probability distributions, topics in linear algebra, scatter matrices, factor analysis, multivariate analysis of variance, and graph theory.

Chapter 3, the main chapter of the book, is also its main strength: the authors do an excellent job in explaining the various algorithms. Actually I was first a bit puzzled by the way single linkage and complete linkage are introduced: These two algorithms are first explained in graph-theoretic terms, and the familiar characterization in terms of minimum and maximum distance follows later. However, at a more careful reading, I found this approach convincing and very elegant.

Those parts of the book which deal with statistical concepts are in general not as good. For instance, Section 2.4.3 talks about discriminant analysis, using the notions of within-groups and between-groups covariance matrices, and Wilks'

Lambda statistic. These notions are explained in appendices, but I doubt whether the statistically untrained reader will be able to follow the presentation. Similarly, Section 2.4.1 treats projections based on eigenvectors, but the discussion of principal components is far too short for the non-specialized reader to understand the purpose of this approach. The appendices on factor analysis and multivariate analysis of variance are so short that they are hardly useful at all. Moreover, statistical terms are not always used properly – for instance, the word 'significant' is used in a very vague sense (p. 95), and the notion of a statistical hypothesis remains rather unclear (p. 144).

Cluster analysis is a wide area, and whoever attempts to write a book on these topics has to face difficult dicisions how to select material. In my opinion, the authors have made a good selection, although I would have liked to see more on clustering based on density estimation and mixture decomposition. Also, I would have preferred to find more real-data examples, perhaps at the cost of sacrificing some of the artificial data examples presented throughout the book.

Overall, despite some criticism, my impression is positive, and I recommend the book to all those who are interested in cluster analysis as a practical method of exploring data.

Bernhard Flury
Indiana University