

David Skillicorn (2007): Understanding Complex Datasets: Data Mining with Matrix Decompositions

Chapman and Hall/CRC, xii + 236 pp., US \$ 69,95, € 49,99,
ISBN 978-1584888321

Ricardo Maronna

Published online: 6 June 2008
© Springer-Verlag 2008

This book is an excellent introduction to a set of important data mining tools. It shows many unexpected applications of matrix factorizations. Its elegant style makes reading pleasant. The book is adequate for any reader familiar with linear algebra, but it will be more fruitful for readers with a knowledge of multivariate analysis. Readers interested in the details of the techniques described in the book are guided to the respective references.

An overview of the contents is as follows. Chapter 1 describes the different types of data and aims in data mining, drawing a distinction between business and scientific data sets. Chapter 2 contains a general discussion of matrix decompositions of the form $A = CWF$, and their different interpretations and applications. Chapter 3 deals with the Singular Value Decomposition (SVD), its uses for denoising and clustering, and different applications, in particular to information retrieval and collaborative filtering. Chapter 4 discusses different ways to deal with graph data, their connections to SVD, and different ways to represent a data set as a graph. A very interesting example is the description of the PageRank algorithm used by Google. Chapter 5 deals with the approximation of a data set A by a product CWF where the elements of C and F are 0, 1 or -1 , and its applications, in particular to the simultaneous clustering of rows and columns of A . Chapter 6 shows the usefulness of combining SVD and the Semi Discrete Decomposition by applying them to mineral exploration and protein conformation. Chapter 7 describes Independent Components Analysis and its applications to microarrays and counterterrorism. Chapter 8 deals with decompositions $A = CF$ where C and F have nonnegative elements, and their applications in microarray analysis and mineral exploration. Chapter 9 discusses multi-indexed arrays, the Tucker3 and CP decompositions, and applications. Chapter 10 is a general discussion of the book's material. The Appendix contains Matlab code.

R. Maronna (✉)
University of La Plata, La Plata, Argentina
e-mail: rmaronna@mail.retina.ar