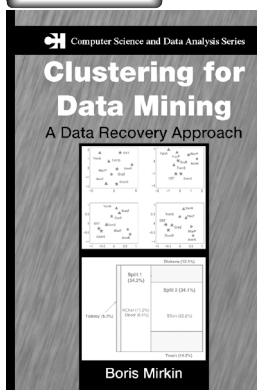


**NEW!** Get more than just *ad hoc* methods—discover a sound theoretical framework!



# Clustering for Data Mining A Data Recovery Approach

**Boris Mirkin**

*School of Computer Science and Information Systems  
Birkbeck College, University of London, UK*

A volume in the **Computer Science and Data Analysis Series**

*Series Editors:* John Lafferty, David Madigan, Fionn Murtagh, and Padhraic Smyth

## DATA RECOVERY EXTENDS TRADITIONAL METHODS TO MODERN TASKS

Often considered more as an art than a science, the field of clustering has been dominated by learning through examples and by choosing techniques almost through trial-and-error. Even the most popular clustering methods — K-Means for partitioning the data set and Ward's method for hierarchical clustering — have lacked the theoretical attention that would establish a firm relationship between the two methods and provide relevant interpretation aids.

Rather than the traditional set of *ad hoc* techniques, **Clustering for Data Mining: A Data Recovery Approach** presents a theory that not only closes gaps in K-Means and Ward methods, but also extends the methods into areas of current interest, such as clustering mixed scale data and incomplete clustering. The author suggests original methods for both cluster finding and cluster description; addresses related topics such as principal component analysis, contingency measures, and data visualization; and includes nearly 60 computational examples covering all stages of clustering, from data pre-processing to cluster validation and results interpretation.

This author's unique attention to data recovery methods, theory-based advice, pre- and post-processing issues and clear, practical instructions for real-world data mining make this book ideally suited for virtually all purposes: for teaching, for self-study, and for professional reference.

### FEATURES

- Introduces classical clustering methods extended, via the data recovery approach, to modern data mining tasks
- Fills gaps in the established theories and corrects common misconceptions
- Treats the two most popular methods, K-Means and Ward clustering, offering the first theoretically motivated instructions for automating all steps of data mining with clustering
- Presents a wealth of computational examples covering all stages of clustering

## CONTENTS

### INTRODUCTION: HISTORICAL REMARKS

#### WHAT IS CLUSTERING

Exemplary Problems  
Bird's Eye View

#### WHAT IS DATA

Feature Characteristics  
Bivariate Analysis  
Feature Space and Data Scatter  
Preprocessing and Standardizing Mixed Data

#### K-MEANS CLUSTERING

Conventional K-Means  
Initialization of K-Means  
Intelligent K-Means  
Interpretation Aids  
Overall Assessment

#### WARD HIERARCHICAL CLUSTERING

Agglomeration: Ward Algorithm  
Divisive Clustering with Ward Criterion  
Conceptual Clustering  
Extensions of Ward Clustering  
Overall Assessment

#### DATA RECOVERY MODELS

Statistics Modeling as Data Recovery  
Data Recovery Model for K-Means  
Data Recovery Models for Ward Criterion

*contents continued on reverse*

**Catalog no. C5343, April 2005, 296 pp.**  
**ISBN: 1-58488-534-3, \$79.95 / £44.99**

 **Chapman & Hall/CRC**  
Taylor & Francis Group

## Contents continued

Extensions to Other Data Types  
One-by-One Clustering  
Overall Assessment

### DIFFERENT CLUSTERING APPROACHES

Extensions of K-Means Clustering  
Graph-Theoretic Approaches  
Conceptual Description of Clusters  
Overall Assessment

### GENERAL ISSUES

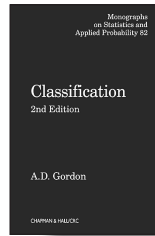
Feature Selection and Extraction  
Data Pre-Processing and Standardization  
Similarity on Subsets and Partitions  
Dealing with Missing Data  
Validity and Reliability  
Overall Assessment

### CONCLUSION: Data Recovery Approach in Clustering

### BIBLIOGRAPHY

Each chapter also contains a section of Base Words

## Other Titles of Interest



### Classification

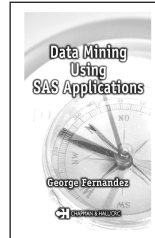
Second Edition

A.D. Gordon

University of Saint Andrews, Scotland

Catalog no. C0346, 1999, 250 pp.

ISBN: 1-58488-013-9, \$89.95 / £44.99



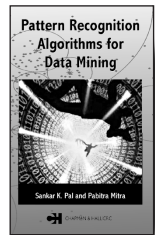
### Data Mining Using SAS Applications

George Fernandez

University of Nevada, Reno, USA

Catalog no. C3456, 2003, 384 pp.

ISBN: 1-58488-345-6, \$79.95 / £39.99



### Pattern Recognition Algorithms for Data Mining

Sankar K. Pal

Indian Statistical Institute, Calcutta, India

Pabitra Mitra

Indian Institute of Technology, Kanpur, India

Catalog no. C4576, 2004, 280 pp.

ISBN: 1-58488-457-6, \$89.95 / £49.99

## Please use this ORDER FORM, CALL or ORDER ONLINE at WWW.CRCPRESS.COM

Please indicate quantities next to the title(s) ordered below:

### CLUSTERING FOR DATA MINING: A DATA RECOVERY APPROACH

.....Catalog no. C5343, ISBN: 1-58488-534-3 at \$79.95 / £44.99 each.

### Other titles of interest:

#### CLASSIFICATION, SECOND EDITION

.....Catalog no. C0346, ISBN: 1-58488-013-9 at \$89.95 / £44.99 each.

#### DATA MINING USING SAS APPLICATIONS

.....Catalog no. C3456, ISBN: 1-58488-345-6 at \$79.95 / £39.99 each.

#### PATTERN RECOGNITION ALGORITHMS FOR DATA MINING

.....Catalog no. C4576, ISBN: 1-58488-457-6 at \$89.95 / £49.99 each.

**Ordering Information:** Orders must be prepaid or accompanied by a purchase order. Checks should be made payable to CRC Press. Please add the appropriate shipping and handling charge for each book ordered. All prices are subject to change without notice. **U.S./Canada:** All orders must be paid in U.S. dollars. **TAX:** As required by law, please add applicable state and local taxes on all merchandise delivered to CA, CT, FL, KY, MO, NY, and PA. For Canadian orders, please add GST. We will add tax on all credit card orders. **European Orders:** All orders must be paid in U.K. £. VAT will be added at the rate applicable. **Textbooks:** Special prices for course adopted textbooks may be available for certain titles. To review a book for class adoption, contact our Academic Sales Department or submit your textbook evaluation request online at [www.crcpress.com/eval.htm](http://www.crcpress.com/eval.htm) **Satisfaction Guaranteed:** If the book supplied does not meet your expectations, it may be returned to us in a saleable condition within 30 days of receipt for a full refund.

### SHIPPING AND HANDLING

Region	Delivery Time	First Title	Additional Title	For priority mail services, please contact your nearest CRC PRESS office.
USA/Canada	3-5 Days	\$5.99	\$1.99	
South America	7-14 Days	\$9.99	\$3.99	
Europe	3-5 Days	£2.99	£0.99	
Rest of World	7-21 Days	£4.99	£2.99	

Name .....  
please print clearly

Company/Institution .....

Address .....

City ..... State/Province ..... Zip/Postal Code .....

Country .....

Visa  MasterCard  American Express  Check Enclosed \$ .....  
Exp. Date: \_\_\_\_\_  
Month Year

Signature and Telephone Number required on all orders

Signature ..... PO# .....

Telephone .....

If you would like to receive information from us by e-mail, please provide your e-mail address below.

E-Mail Address .....

## ORDERING LOCATIONS

### In the Americas: CRC PRESS

6000 Broken Sound Parkway, NW, Suite 300  
Boca Raton, FL 33487, USA  
Tel: 1-800-272-7737  
Fax: 1-800-374-3401  
From Outside the Continental U.S.  
Tel: 1-561-994-0555  
Fax: 1-561-361-6018  
e-mail: [orders@taylorandfrancis.com](mailto:orders@taylorandfrancis.com)

### Rest of the World: CRC PRESS / ITPS

Cheriton House, North Way  
Andover, Hants, SP10 5BE, UK  
Tel (UK): +44 (0) 1264 34 2926  
Tel (Int'l): +44 (0) 1264 34 3070  
Fax: +44 (0) 1264 34 3005  
e-mail:  
(UK): [uk.tandf@thomsonpublishingservices.co.uk](mailto:uk.tandf@thomsonpublishingservices.co.uk)  
(Int'l): [international.tandf@thomsonpublishingservices.co.uk](mailto:international.tandf@thomsonpublishingservices.co.uk)

## Corporate Offices

### CRC PRESS

6000 Broken Sound Parkway, NW, Suite 300  
Boca Raton, FL 33487, USA  
Tel: 1-800-272-7737  
Fax: 1-800-374-3401  
From Outside the Continental U.S.  
Tel: 1-561-994-0555  
Fax: 1-561-361-6018  
e-mail: [orders@taylorandfrancis.com](mailto:orders@taylorandfrancis.com)

### CRC PRESS UK

23-25 Blades Court, Deodar Road  
London SW15 2NU, UK  
Tel: 44 (0) 20 8875 4370  
Fax: 44 (0) 20 8871 3443  
e-mail: [enquiries@crcpress.com](mailto:enquiries@crcpress.com)

[www.crcpress.com](http://www.crcpress.com)

3.29.05mn