

Henk A.L. Kiers · Jean-Paul Rasson
Patrick J.F. Groenen · Martin Schader (Eds.)

Data Analysis, Classification, and Related Methods

With 96 Figures



Springer

Contents

Part I. Cluster Analysis

Cluster Analysis and Mixture Models

Classifier Probabilities	3
<i>J. A. Hartigan</i>	
Cluster Analysis Based on Data Depth	17
<i>Richard Hoberg</i>	
An Autonomous Clustering Technique	23
<i>Yoshiharu Sato</i>	
Unsupervised Non-hierarchical Entropy-based Clustering	29
<i>M. Jardino</i>	
Improving the Additive Tree Representation of a Dissimilarity Matrix Using Reticulations	35
<i>Vladimir Makarenkov, Pierre Legendre</i>	
Double Versus Optimal Grade Clusterings	41
<i>Alicja Ciok</i>	
The Effects of Initial Values and the Covariance Structure on the Recovery of some Clustering Methods	47
<i>Istvan Hajnal, Geert Loosveldt</i>	
What Clusters Are Generated by Normal Mixtures?.....	53
<i>Christian Hennig</i>	
A Bootstrap Procedure for Mixture Models	59
<i>Suzanne Winsberg, Geert deSoete</i>	
<i>Fuzzy Clustering</i>	
A New Criterion of Classes Validity	63
<i>Arnaud Devillez, Patrice Billaudel, Gérard Villerman Lecolier</i>	
Application of Fuzzy Mathematical Morphology for Unsupervised Color Pixels Classification	69
<i>A. Gillet, C. Botte-Lecocq, L. Macaire and J.-G. Postaire</i>	

A Hyperbolic Fuzzy k-Means Clustering and Algorithm for Neural Networks	77
<i>Norio Watanabe, Tadashi Imaizumi, Toshiko Kikuchi</i>	
<i>Special Purpose Classification Procedures and Applications</i>	
A Toolkit for Development of the Domain-Oriented Dictionaries for Structuring Document Flows	83
<i>Pavel P. Makagonov, Mikhail A. Alexandrov, Konstantin Sboychakov</i>	
Classification of Single Malt Whiskies	89
<i>David Wishart</i>	
Robust Approach in Hierarchical Clustering: Application to the Sectorisation of an Oil Field	95
<i>Jean-Paul Valois</i>	
A Minimax Solution for Sequential Classification Problems ...	101
<i>Hans J. Vos</i>	
<i>Verification and Comparison of Clusterings</i>	
Comparison of Ultrametries Obtained With Real Data, Using the P_L and VAL_{Aw} Coefficients	107
<i>Isabel Pinto Doria, Georges Le Calvé, Helena Bacelar-Nicolau</i>	
Numerical Comparisons of two Spectral Decompositions for Vertex Clustering	113
<i>P. Kuntz, F. Henaux</i>	
Measures to Evaluate Rankings of Classification Algorithms	119
<i>Carlos Soares, Pavel Brazdil, Joaquim Costa</i>	
A General Approach to Test the Pertinence of a Consensus Classification	125
<i>Guy Cucumel, François-Joseph Lapointe</i>	
<i>Dissimilarity Measures</i>	
On a Class of Aggregation-invariant Dissimilarities Obeying the Weak Huygens' Principle	131
<i>F. Bavaud</i>	
A Short Optimal Way for Constructing Quasi-ultrametrics From Some Particular Dissimilarities	137
<i>B. Fichet</i>	

Missing Data in Cluster Analysis

Estimating Missing Values in a Tree Distance	143
<i>A. Guénoche, S. Grandcolas</i>	
Estimating Trees From Incomplete Distance Matrices: A Comparison of Two Methods	149
<i>Claudine Levasseur, Pierre-Alexandre Landry, François-Joseph Lapointe</i>	
Zero Replacement in Compositional Data Sets	155
<i>J. A. Martín-Fernández, C. Barceló-Vidal, V. Pawlowsky-Glahn</i>	
EM Algorithm for Partially Known Labels	161
<i>C. Ambroise, G. Govaert</i>	

Part II. Discrimination, Regression Trees, and Data Mining

Discriminant Analysis

Detection of Company Failure and Global Risk Forecasting ...	169
<i>Mireille Bardos</i>	
Discriminant Analysis by Hierarchical Coupling in EDDA Context	175
<i>Isabel Brito, Gilles Celeux</i>	
Discrete Discriminant Analysis: The Performance of Combining Models by a Hierarchical Coupling Approach ...	181
<i>Ana Sousa Ferreira, Gilles Celeux, Helena Bacelar-Nicolau</i>	
Discrimination Based on the Atypicality Index versus Density Function Ratio	187
<i>H. Chamlal and S. Slaoui Chah</i>	

Decision and Regression Trees

A Third Stage in Regression Tree Growing: Searching for Statistical Reliability	193
<i>Carmela Cappelli, Francesco Mola, Roberta Siciliano</i>	
A New Sampling Strategy for Building Decision Trees from Large Databases	199
<i>J.H. Chauchat, R. Rakotomalala</i>	

Generalized Additive Multi-Model for Classification and Prediction	205
<i>Claudio Conversano, Francesco Mola, Roberta Siciliano</i>	
Radial Basis Function Networks and Decision Trees in the Determination of a Classifier	211
<i>Rossella Miglio, Marilena Pillati</i>	
Clustered Multiple Regression	217
<i>Luis Torgo, J. Pinto da Costa</i>	
<i>Neural Networks and Data Mining</i>	
Artificial Neural Networks, Censored Survival Data, Statistical Models	223
<i>Antonio Ciampi, Yves Lechevallier</i>	
Visualisation and Classification with Artificial Life	229
<i>Alfred Ultsch</i>	
<i>Pattern Recognition and Geometrical Statistics</i>	
Exploring the Periphery of Data Scatters: Are There Outliers?	235
<i>Giovanni C. Porzio, Giancarlo Ragozini</i>	
Discriminant Analysis Tools for Non Convex Pattern Recognition	241
<i>Marcel Rémon</i>	
A Markovian Approach to Unsupervised Multidimensional Pattern Classification	247
<i>A. Sbihi, A. Moussa, B. Benmiloud, J.-G. Postaire</i>	

Part III. Multivariate and Multidimensional Data Analysis

Multivariate Data Analysis

An Algorithm with Projection Pursuit for Sliced Inverse Regression Model	255
<i>Masahiro Mizuta, Hiroyuki Minami</i>	
Testing Constraints and Misspecification in VAR-ARCH Models	261
<i>Wolfgang Polasek, Shuangzhe Liu</i>	

Goodness of Fit Measure based on Sample Isotone Regression of Mokken Double Monotonicity Model	267
<i>Teresa Rivas Moya</i>	

Multiway Data Analysis

Fuzzy Time Arrays and Dissimilarity Measures for Fuzzy Time Trajectories	273
<i>Renato Coppi, Pierpaolo D'Urso</i>	

Three-Way Partial Correlation Measures	279
<i>Donatella Vicari</i>	

Analysis of Network and Relationship Data and Multidimensional Scaling

Statistical Models for Social Networks	285
<i>Stanley Wasserman, Philippa Pattison</i>	

Application of Simulated Annealing in some Multidimensional Scaling Problems	297
<i>Javier Trejos, William Castillo, Jorge González, Mario Villalobos</i>	

Data Analysis Based on Minimal Closed Subsets	303
<i>S. Bonnevey, C. Largeron-Leteno</i>	

Robust Multivariate Methods

A Robust Method for Multivariate Regression	309
<i>Stefan Van Aelst, Katrien Van Driessen, Peter J. Rousseeuw</i>	

Robust Methods for Complex Data Structures	315
<i>Ursula Gather, Claudia Becker, Sonja Kuhnt</i>	

Robust Methods for Canonical Correlation Analysis	321
<i>Catherine Dehon, Peter Filzmoser, Christophe Croux</i>	

Part IV. Data Science

Data Science and Data Collection

From Data Analysis to Data Science	329
<i>Noboru Ohsumi</i>	

Evaluation of Data Quality and Data Analysis	335
<i>Chikio Hayashi</i>	

**Collapsibility and Collapsing Multidimensional Contingency
Tables—Perspectives and Implications** 341
Stefano De Cantis, Antonino M. Oliveri

Sampling and Internet Surveys

Data Collected on the Web 347
Vasja Vehovar, Katja Lozar Manfreda, Zenel Batagelj

**Some Experimental Surveys on the WWW Environments
in Japan** 353
Osamu Yoshimura, Noboru Ohsumi

**Bootstrap Goodness-of-fit Tests
for Complex Survey Samples** 359
Andrea Scagni

Part V. Symbolic Data Analysis

Classification and Analysis of Symbolic Data

Regression Analysis for Interval-Valued Data 369
L. Billard, E. Diday

Symbolic Approach to Classify Large Data Sets..... 375
*Francisco de A.T. de Carvalho, Cezar A. de F. Anselmo, Renata
M.C.R. de Souza*

**Factorial Methods with Cohesion Constraints
on Symbolic Objects**..... 381
N.C. Lauro, R. Verde, F. Palumbo

A Dynamical Clustering Algorithm for Multi-nominal Data... 387
Rosanna Verde, Francisco de A. T. de Carvalho, Yves Lechevallier

Software

**DB2SO : A Software for Building Symbolic Objects
from Databases** 395
Georges Hébrail, Yves Lechevallier

**Symbolic Data Analysis and the SODAS Software
in Official Statistics** 401
Raymond Bisdorff, Edwin Diday

Strata Decision Tree SDA Software 409
M. Carmen Bravo

**Marking and Generalization by Symbolic Objects
in the Symbolic Official Data Analysis Software.....** 417
 Mireille Gettler Summa

List of Reviewers 423

Index 425