



17<sup>th</sup> Conference of the  
International Federation  
of Classification Societies

## Classification and Data Science in the Digital Age

# Book of Abstracts



---

---

IFCS 2022

Book of Abstracts

17th Conference of the International Federation of  
Classification Societies

Classification and Data Science  
in the Digital Age

---

PORTO, PORTUGAL

---

---

Title: Classification and Data Science in the Digital Age - Book of Abstracts IFCS  
2022  
Authors: CLAD - Associação Portuguesa de Classificação e Análise de Dados  
Cover design: exclamação!  
Printed in Portugal by Instituto Nacional de Estatística  
ISBN: 978-989-98955-9-1  
N. DL: 500335/22  
Number of copies: 350

# Preface

Welcome to the 17<sup>th</sup> conference of the IFCS, IFCS 2022, held in Porto, Portugal, from July 19 to July 23, 2022 and the first IFCS conference held in Portugal. It is a joint organisation of the Portuguese Association for Classification and Data Analysis, CLAD, and the Faculty of Economics of the University of Porto, FEP-UP.

IFCS 2022 is preceded by two half-day tutorials, one on *Analysis of Data Streams* by João Gama, and another on *Categorical Data Analysis and Visualization* by Rosaria Lombardo and Eric J Beh, features four keynote speakers, five invited and seventy contributed sessions organised in specific topics. The Benchmarking Challenge, the Awards Session and the Presidential Address are also noteworthy. Overall, the call for papers attracted 280 submissions, representing 42 countries and 578 different authors. The authors come from five continents, being the largest representation from Europe (68%), followed by North America (12%). Additionally to the rich scientific program the LOC has organised a number of social appealing events that will be memorable.

The 17<sup>th</sup> conference of the IFCS would not have been possible without the support of many individuals and organisations. We owe special thanks to the authors of all the submitted papers, the members of the program committee, and the reviewers for their contributions to the success of the conference. Finally, we acknowledge the institutional and industrial sponsors that contributed to the organisation of the conference. In particular, we thank all those at FEP-UP who enthusiastically supported the conference from the very start, contributing to its success.

This book contains the abstracts corresponding to all the presentations at the conference. It is organised in seven parts, according to the type of session. Within each session the abstracts are ordered according to the programme. The book includes also an author index.

It has been a pleasure and an honor to organise and host IFCS 2022 in Porto. It is our wish that all participants enjoy the scientific program as well as the the social events and the city of Porto and Portugal.

July 2022

The Local Organising Committee

# Contents

## Part I Presidential Address

<b>Perturb and Conquer: how Classification Can Benefit from Data Perturbation</b> . . . . .	2
Angela Montanari	

## Part II IFCS Research Medal Lecture

<b>An Introduction to S-concordance and S-discordance</b> . . . . .	4
Edwin Diday	

## Part III Tutorials

<b>Analysis of Data Streams</b> . . . . .	6
João Gama	
<b>Categorical Data Analysis and Visualization</b> . . . . .	7
Rosaria Lombardo and Eric J. Beh	

## Part IV Keynote Lectures

<b>A Showcase of New Methods for High Dimensional Data Viewing with Linear Projections and Sections</b> . . . . .	10
Dianne Cook	
<b>Statistical Learning with Dynamic Interaction Data for Public Health</b> . . .	11
Charles Bouveyron	
<b>Fast Minipatch Ensemble Strategies for Learning and Inference</b> . . . . .	12
Genevera I. Allen	
<b>Trends in Data Stream Mining</b> . . . . .	13
João Gama	

**Part V Invited Sessions**

<b>A Criterion for Selecting the Number of Time Series Clusters</b> . . . . .	16
Daniel Peña and Ruey S. Tsay	
<b>Clusters Based on Prediction Accuracy of Global Forecasting Models</b> . . .	17
Pablo Montero-Manso, Ángel López-Oriona, and José A. Vilar	
<b>Clustering and Classifying Time Series in the Sktime Toolkit: a Practical Review of Latest Advances in the Field</b> . . . . .	18
Anthony Bagnall	
<b>Effect of Type 2 Diabetes and its Genetic Susceptibility on Severity and Mortality of COVID-19 in UK Biobank</b> . . . . .	19
Aeyeon Lee, Youngkwang Cho, Jun Li, Taesung Park, Wonil Chung, and Liming Liang	
<b>Prognosis of COVID-19 Patients by the Underlying Diseases and Drug Treatment in Korea</b> . . . . .	20
Ho Kim and Taerim Lee	
<b>Algorithms for Clustering COVID-19 Data: a Holistic Overview of Current Trends and New Visual Approaches</b> . . . . .	21
Eun-Kyung Lee	
<b>Embedded Word MCA Biplots for Sentiment Visualisation: Application to COVID-19 Related Tweets</b> . . . . .	22
Zoë-Mae Adams, Johané Nienkemper-Swanepoel, Niël le Roux, and Sugnet Lubbe	
<b>A General Framework for Implementing Distance Measures for Categorical Variables</b> . . . . .	23
Carlo Cavicchia, Michel van de Velden, Alfonso Iodice D’Enza, and Angelos Markos	
<b>Some Descriptive Statistics of Aggregated Symbolic Data</b> . . . . .	24
Junji Nakano, Nobuo Shimizu, and Yoshikazu Yamamoto	
<b>Variable Screening in High Dimensional Regression via Random Projection Ensembles</b> . . . . .	25
Laura Anderlucchi, Matteo Farnè, Giuliano Galimberti, and Angela Montanari	
<b>Model-based Clustering and Dimension Reduction for Multidimensional Social Networks</b> . . . . .	26
Michael Fop, Silvia D’Angelo, and Marco Alfò	
<b>Conditional Gaussian Mixture Modeling</b> . . . . .	27
Volodymyr Melnykov and Yang Wang	

<b>Classification Over Text, Relational Databases and Graphs - Software and Case Studies</b> .....	28
Tomáš Kliegr	
<b>Towards Deep and Interpretable Rule Learning</b> .....	29
Johannes Fürnkranz	
<b>Current Challenges in Interpretable Machine Learning and Partitioning Approaches</b> .....	30
Bernd Bischl	
<b>Part VI Benchmarking Challenge</b>	
<b>Vine Copula Mixture Models and Clustering for Non-Gaussian Data</b> . . . .	32
Özge Sahin and Claudia Czado	
<b>Comparison of Similarity Measures for Categorical Data in Hierarchical Clustering</b> .....	33
Zdeněk Šulc and Hana Řezanková	
<b>Comparing Model Selection Techniques to Determine the Number of Overlapping Clusters for the Additive Profile Clustering Model</b> .....	34
Tom F. Wilderjans, Julian Rossbroich, and Jeffrey Durieux	
<b>Pitfalls of Automatic Optimization Procedures and Benchmarking in Cluster Analysis</b> .....	35
Quirin Stier and Michael C. Thrun	
<b>Part VII Contributed Abstracts</b>	
<b>Biplots for Categorical and Ordinal Data Based on Logistic Responses</b> . .	38
Jose Luis Vicente-Villardón	
<b>The Biplot Inner Product for Interpretation and Derivation of Eigenvector Methods</b> .....	39
Cajo J. F. ter Braak	
<b>Fifty Years of Biplots: Some Remaining Enigmas and Challenges</b> .....	40
Jan Graffelman	
<b>Outlier Detection for BIG Functional Data</b> .....	41
Rosa E. Lillo, Oluwasegun T. Ojo, and Antonio Fernández-Anta	
<b>Outlier and Novelty Detection for Functional Data: a Semiparametric Bayesian Approach</b> .....	42
Francesco Denti, Andrea Cappozzo, and Francesca Greselin	
<b>A Geometric Perspective on Functional Outlier Detection</b> .....	43
Moritz Herrmann and Fabian Scheipl	

Contents	xix
<b>A New Decomposition of Orthogonal Matrices with Application to Common Principal Components</b> .....	44
Luca Bagnato and Antonio Punzo	
<b>An MML Embedded Approach for Estimating the Number of Clusters</b> ..	45
Cláudia Silvestre, Margarida G. M. S. Cardoso, and Mário Figueiredo	
<b>Comparison of Segmentation Approaches for Partial Least Squares Path Modeling with Stability Assessment</b> .....	46
Sophie Dominique, Mohamed Hanafi, Fabien Llobell, Jean-Marc Ferrandi, and Véronique Cariou	
<b>Robust Classification for Toroidal Data</b> .....	47
Giovanni Saraceno, Luca Greco, and Claudio Agostinelli	
<b>Consistency of Trimmed Estimators of Scatter Under the t-distribution</b> ..	48
Andrea Cerioli, Lucio Barabesi, Luis A. García-Escudero, and Agustín Mayo-Iscar	
<b>Robust Classification in High Dimensions Using Regularized Covariance Estimates</b> .....	49
Valentin Todorov and Peter Filzmoser	
<b>Symbolic Concordance and Discordance Illustrated on Data from an International Teaching and Learning Survey</b> .....	50
Simona Korenjak-Černe, Barbara Japelj Pavešić, and Edwin Diday	
<b>A Clusterwise Regression Method for Distributional Data</b> .....	51
Rosanna Verde, Antonio Balzanella, and Antonio Irpino	
<b>The Use of Regression to Partition a Dataset of Interval Observations</b> ...	52
Lynne Billard and Fei Liu	
<b>Heterogeneous Random Forests</b> .....	53
Ye-eun Kim and Hyunjoong Kim	
<b>Analysis of the Damage Rate Using Typhoon Information</b> .....	54
Su Hoon Choi and Min Soo Kim	
<b>Resampling, Relabeling, and Raking for Extremely Imbalanced Classification</b> .....	55
Hae-Hwan Lee, Seunghwan Park, and Jongho Im	
<b>Clustering Student Mobility Data in 3-way Networks</b> .....	56
Vincenzo Giuseppe Genova, Giuseppe Giordano, Giancarlo Ragozini, and Maria Prosperina Vitale	
<b>Multi-perspective Risky User Classification in Social Networks</b> .....	57
Antonio Pellicani, Gianvito Pio, and Michelangelo Ceci	



<b>Clustering and Blockmodeling Temporal Networks – Two Indirect Approaches</b> . . . . .	58
Vladimir Batagelj	
<b>Clustering Validation in Hierarchical Cluster Analysis: an Empirical Study</b> . . . . .	59
Osvaldo Silva, Áurea Sousa, and Helena Bacelar-Nicolau	
<b>Divide and Conquer: a Clustering Method for Hierarchical and Nested Data Structures</b> . . . . .	60
Andrej Svetlošák, Miguel de Carvalho, Gabriel Martos Venturini, and Raffaella Calabrese	
<b>Significance Mode Analysis (SigMA) for Hierarchical Structures</b> . . . . .	61
Sebastian Ratzenböck, Torsten Möller, Josefa E. Großschedl, João Alves, Immanuel M. Bomze, and Stefan Meingast	
<b>Kurtosis-based Projection Pursuit for Matrix-valued Data</b> . . . . .	62
Una Radojicic, Klaus Nordhausen, and Joni Virta	
<b>Comparison of Pixel Based Segmentation Methods in Papillary Thyroid US Images</b> . . . . .	63
Neslihan Gökmen İnan, İsmail Meşe, Düzgün Yıldırım, and Ozan Kocadağlı	
<b>Bootstrapping Binary GEV Regressions for Massive Unbalanced Datasets</b> . . . . .	64
Michele La Rocca, Marcella Niglio, and Marialuisa Restaino	
<b>Stochastic Collapsed Variational Inference for Structured Gaussian Process Regression Networks</b> . . . . .	65
Rui Meng, Herbert K. H. Lee, and Kristofer Bouchard	
<b>Covariate Selection Method in Propensity Score Model for the Quantile Treatment Effect Estimation</b> . . . . .	66
Takehiro Shoji, Jun Tsuchida, and Hiroshi Yadohisa	
<b>Are Attitudes Toward Immigration Changing in Europe? An Analysis Based on Latent Class IRT Models</b> . . . . .	67
Ewa Genge and Francesco Bartolucci	
<b>Visualization of IATA Regions in Air Transport Before and After the COVID-19 Pandemic</b> . . . . .	68
Tüzün Tolga İnan, Neslihan Gökmen İnan, Aylin Yaman Kocadağlı, and Ozan Kocadağlı	
<b>Political and Religion Attitudes in Greece: Behavioral Discourses</b> . . . . .	69
Georgia Panagiotidou and Theodore Chadjipadelis	
<b>Functional Data Representation with Merge Trees</b> . . . . .	70
Matteo Pegoraro and Piercesare Secchi	

Contents	xxi
<b>Elastic Regression for Irregularly Sampled Curves in <math>\mathbb{R}^d</math></b> . . . . .	71
Lisa Steyer, Almond Stöcker, and Sonja Greven	
<b>Misalignment of Spectral Data: Constrained Optimization in a Functional Data Analysis Framework</b> . . . . .	72
Francesca Di Salvo, Delia Francesca Chillura Martino, and Gabriella Chirco	
<b>Model Based Clustering of Functional Data with Mild Outliers</b> . . . . .	73
Cristina Anton and Iain Smith	
<b>Old and New Constraints in Model Based Clustering</b> . . . . .	74
Luis A. García-Escudero, Agustín Mayo-Iscar, Gianluca Morelli, and Marco Riani	
<b>Model Based Clustering and Outlier Detection with Missing Data</b> . . . . .	75
Cristina Tortora, Hung Tong, and Louis Tran	
<b>How to Mitigate the Effect of Outliers on Balancing Technique</b> . . . . .	76
Rasool Taban, Maria do Rosário Oliveira, and Claudia Nunes Philippart	
<b>Outliers Detection in Functional Data</b> . . . . .	77
Amovin-Assagba Martial, Gannaz Irène, and Jacques Julien	
<b>Robustified Elastic Net Estimator for Multinomial Regression</b> . . . . .	78
Fatma Sevinç Kurnaz and Peter Filzmoser	
<b>Optimized Symbolic Correspondence Analysis for Multi-valued Variables</b> . . . . .	79
Jorge Arce Garro and Oldemar Rodríguez Rojas	
<b>Symbolic t-SNE and UMAP Methods for Interval Type Variables</b> . . . . .	80
Oldemar Rodríguez Rojas	
<b>Two-stage Principal Component Analysis on Interval-valued Data Using Patterned Covariance Structure</b> . . . . .	81
Anuradha Roy	
<b>Detection of the Biliary Atresia Using Deep Convolutional Neural Networks Based on Statistical Learning Weights via Optimal Similarity and Resampling Methods</b> . . . . .	82
Kuniyoshi Hayashi, Eri Hoshino, Mitsuyoshi Suzuki, Erika Nakanishi, Kotomi Sakai, and Masayuki Obatake	
<b>Variational Autoencoder with Gamma Mixture for Clustering Right-skewed Data</b> . . . . .	83
Jinwon Heo and Jangsun Baek	
<b>An Efficient Way to Identify Inliers via Inlier-memorization Effect of Deep Generative Models</b> . . . . .	84
Dongha Kim, Jaesung Hwang, and Yongdai Kim	

<b>Three-way Spectral Clustering</b> . . . . .	85
Cinzia Di Nuzzo and Salvatore Ingrassia	
<b>Fuzzy Clustering by Hyperbolic Smoothing</b> . . . . .	86
David Masís, Esteban Segura, Javier Trejos, and Adilson Xavier	
<b>Combining KDE and DBSCAN Clustering to Understand Road Traffic Accidents: the Case of Setúbal, Portugal</b> . . . . .	87
Pedro Nogueira, Marcelo Silva, Paulo Infante, Paulo Rebelo Manuel, Leonor Rego, Anabela Afonso, and Gonçalo Jacinto	
<b>Similarity Forest for Time Series Classification</b> . . . . .	88
Tomasz Górecki, Maciej Łuczak, and Paweł Piasecki	
<b>Uncovering Regions of Maximum Dissimilarity on Random Process Data</b>	89
Miguel de Carvalho and Gabriel Martos Venturini	
<b>Franz Liszt's Transcendental Études: an Evolutionary Analysis by Machine Learning</b> . . . . .	90
Matteo Farnè	
<b>Quantile-distribution Functions and Their Use for Classification</b> . . . . .	91
Edoardo Redivo, Cinzia Viroli, and Alessio Farcomeni	
<b>Analysis of Gini Splitting Criterion and Comparison with Maximum Likelihood Rule</b> . . . . .	92
Amirah S. Alharthi and Charles C. Taylor	
<b>Envelope-based Support Vector Machine Classifier</b> . . . . .	93
Alya Alzahrani and Andreas Artemiou	
<b>A Moment-free Measure of Multivariate Skewness</b> . . . . .	94
Andrzej Sokolowski and Malgorzata Markowska	
<b>The Weighted RV Coefficient: Exact Moments by Invariant Orthogonal Integration</b> . . . . .	95
François Bavaud	
<b>Testing Equality of Multivariate Coefficients of Variation</b> . . . . .	96
Marc Ditzhaus and Łukasz Smaga	
<b>A New Regression Model for the Analysis of Microbiome Data</b> . . . . .	97
Roberto Ascari and Sonia Migliorati	
<b>The Death Process in Italy Before and During the Covid-19 Pandemic: a Functional Compositional Approach</b> . . . . .	98
Riccardo Scimone, Alessandra Menafoglio, Laura M. Sangalli, and Piercesare Secchi	

<b>Sampling Design for Uncovering Natural Laws in Compositional Data . . .</b>	99
Lan Liang, Glòria Mateu-Figueras, and Jan Graffelman	
<b>Penalized Model-based Functional Clustering: a Regularization Approach via Shrinkage Methods . . . . .</b>	100
Nicola Pronello, Rosaria Ignaccolo, Luigi Ippoliti, and Sara Fontanella	
<b>Clustering in FDA Mixing the Epigraph and the Hypograph Indexes with Machine Learning Algorithms . . . . .</b>	101
Belén Pulido, Alba M. Franco-Pereira, and Rosa E. Lillo	
<b>Localization Processes for Functional Data Classification . . . . .</b>	102
Antonio Elías, Raúl Jiménez, and Joseph E. Yukich	
<b>A New Functional Data Clustering Technique Based on Spectral Clustering and Downsampling . . . . .</b>	103
Maryam Al Alawi, Surajit Ray, and Mayetri Gupta	
<b>Parsimonious Mixtures of Seemingly Unrelated Contaminated Normal Regression Models . . . . .</b>	104
Gabriele Perrone and Gabriele Soffritti	
<b>Monitoring Hyperparameter Choice for Robust Cluster Weighted Model</b>	105
Andrea Cappozzo, Luis A. García-Escudero, Francesca Greselin, and Agustín Mayo-Iscar	
<b>Latent Block Regression Model . . . . .</b>	106
Rafika Boutalbi, Lazhar Labiod, and Mohamed Nadif	
<b>Towards a Bi-stochastic Matrix Approximation of <math>k</math>-means and Some Variants . . . . .</b>	107
Lazhar Labiod and Mohamed Nadif	
<b>Clustering Brain Connectomes Through a Density-peak Approach . . . . .</b>	108
Riccardo Giubilei	
<b>New Metrics for Classifying Phylogenetic Trees Using <math>k</math>-means and the Symmetric Difference Metric . . . . .</b>	109
Nadia Tahiri and Aleksandr Koshkarov	
<b>Alternating Optimization Framework for Sparse Simultaneous Component Analysis Based on Data Integration . . . . .</b>	110
Rosember Guerra-Urzola, Juan C. Vera, Katrijn Van Deun, and Klaas Sijtsma	
<b>Joint Sparse Principal Component Analysis . . . . .</b>	111
Katrijn Van Deun	
<b>Joint Sparse Principal Component Analysis: a Simulation Study . . . . .</b>	112
Tra Le and Katrijn Van Deun	

<b>Copula-based Non-metric Unfolding on Augmented Data Matrix</b> . . . . .	113
Marta Nai Ruscone and Antonio D'Ambrosio	
<b>Emotion Classification Based on Single Electrode Brain Data: Applications for Assistive Technology</b> . . . . .	114
Duarte Rodrigues, Luis Paulo Reis, and Brígida Mónica Faria	
<b>On the Role of Data, Statistics and Decisions in a Pandemic</b> . . . . .	115
Ursula Garczarek, Beate Jahn, Sarah Friedrich, Joachim Behnke, Joachim Engel, Ralf Münnich, Markus Pauly, Adalbert Wilhelm, Olaf Wolkenhauer, Markus Zwick, Uwe Sieber, and Tim Friede	
<b>A Deep Learning Analytics to Detect Dental Caries</b> . . . . .	116
Taerim Lee	
<b>Identification of Shared Genetic Loci Between Psychiatric Disorders and Telomere Length and Evaluation of Their Role as Potential Drug Targets</b>	117
Claudia Pisanu, Anna Meloni, and Alessio Squassina	
<b>Estimating Optimal Decision Trees for Treatment Assignment with <math>k &gt; 2</math> Treatment Alternatives: a Classification Problem with a Unit- and Class- dependent Misclassification Cost</b> . . . . .	118
Iven Van Mechelen and Aniek Sies	
<b>ExactTree: an R-package for Globally Optimal Decision Trees</b> . . . . .	119
Elise Dusseldorp, Juan Claramunt Gonzales, Jacqueline Meulman, Samil Uysal, and Bart Jan van Os	
<b>Optimal Random Projection Trees Ensemble</b> . . . . .	120
Nosheen Faiz, Adi Lausen, Metodi Metodiev, Zardad Khan, and Berthold Lausen	
<b>Born-again and Bayesian Approaches for Improving the Performance of Decision Trees</b> . . . . .	121
Marjolein Fokkema	
<b>Evolution of Media Coverage on Climate Change and Environmental Awareness: an Analysis of Tweets from UK and US Newspapers</b> . . . . .	122
Gianpaolo Zammarchi, Maurizio Romano, and Claudio Conversano	
<b>Improving Classification of Documents by Semi-supervised Clustering in a Semantic Space</b> . . . . .	123
Jasminka Dobša and Henk A.L. Kiers	
<b>Is It Hate or Criticism? An Exploratory Approach to Negative Comments on YouTube</b> . . . . .	124
Manuela Schmidt	

<b>A Time-varying Text Based Ideal Point Model to Infer Partisanship in the U.S. Senate</b> . . . . .	125
Sourav Adhikari, Bettina Grün, and Paul Hofmarcher	
<b>Oracle-LSTM: a Neural Network Approach to Mixed Frequency Time Series Prediction</b> . . . . .	126
Alessandro Bitetto and Paola Cerchiello	
<b>Time Series of Counts Under Censoring</b> . . . . .	127
Isabel Silva, Maria Eduarda Silva, Isabel Pereira, and Brendan McCabe	
<b>Multivariate Time Series Feature Extraction via Multilayer Networks</b> . . . . .	128
Vanessa Freitas Silva, Maria Eduarda Silva, Pedro Ribeiro, and Fernando Silva	
<b>On the Use of the Choquet Fuzzy Integral to Aggregate Predictions of Time Series: an Application to Economic (and Other Types of) Data</b> . . . . .	129
Diogo Alves, José Matos, and Sandra Silva	
<b>Some Biplot Alternatives</b> . . . . .	130
Patrick Groenen	
<b>Biplots in Dimension Reduction and Clustering</b> . . . . .	131
Alfonso Iodice D’Enza, Angelos Markos, and Michel van de Velden	
<b>Biplots: a Sophisticated Multivariate Approach or a User-Friendly Technique?</b> . . . . .	132
Manuel Rui Alves	
<b>PLS-based Principal Balances for Regression and Classification with High-dimensional Compositional Data</b> . . . . .	133
Viktorie Nesrstová, Ines Wilms, Karel Hron, Josep A. Martín-Fernández, Peter Filzmoser, and Javier Palarea-Albaladejo	
<b>Clustering Count Data Using Compositional Methods</b> . . . . .	134
Marc Comas-Cufí, Josep A. Martín-Fernández, Glària Mateu-Figueras, and Javier Palarea-Albaladejo	
<b>Urban Development Paths in Poland: Multidimensional Perspective</b> . . . . .	135
Barbara Batóg and Jacek Batóg	
<b>Analyzing the Evolution of EU Countries and Indicators of Europe 2020 Agenda</b> . . . . .	136
Adelaide Figueiredo and Fernanda Figueiredo	
<b>Google Trends as a Macroeconomic Predictor: Behind the Scenes</b> . . . . .	137
Eduardo Andre Costa and Maria Eduarda Silva	

<b>COVID-19 Pandemic: a Methodological Model for the Analysis of Government Preventing Measures and Health Data Records</b> . . . . .	138
Theodore Chadjipadelis and Sofia Magopoulou	
<b>Detecting Fabricated Interviews Using the Hamming Distance</b> . . . . .	139
Joerg Blasius	
<b>Digital Development and Internet Use in the European Union Countries</b> .	140
Fernanda Figueiredo and Adelaide Figueiredo	
<b>Probabilistic Clustering with Local Aligement of Italian COVID-19 Death Curves</b> . . . . .	141
Marzia A. Cremona, Tobia Boschi, and Francesca Chiaromonte	
<b>Model Free Predictive Inference for Functional Kriging Techniques Based on Conformal Prediction</b> . . . . .	142
Andrea Diana, Elvira Romano, and Jorge Mateu	
<b>Density Modelling via Functional Data Analysis</b> . . . . .	143
Stefano A. Gattone and Tonio Di Battista	
<b>On Parsimonious Modelling via Matrix-variate t Mixtures</b> . . . . .	144
Salvatore D. Tomarchio	
<b>Four Skewed Tensor Variate Distributions</b> . . . . .	145
Michael P.B. Gallagher, Peter A. Tait, and Paul D. McNicholas	
<b>A Family of Skewed Power Exponential Mixture Models for Clustering and Classification</b> . . . . .	146
Utkarsh J. Dang, Michael P. B. Gallagher, Ryan P. Browne, and Paul D. McNicholas	
<b>Generating Collective Counterfactual Explanations in Score-based Classification via Mathematical Optimization</b> . . . . .	147
Jasone Ramírez-Ayerbe, Emilio Carrizosa, and Dolores Romero Morales	
<b>Spherical Separation in Machine Learning</b> . . . . .	148
Matteo Avolio, Annabella Astorino, and Antonio Fuduli	
<b>Model Extraction Based on Counterfactual Explanations</b> . . . . .	149
Cecilia Salvatore and Veronica Piccialli	
<b>Isolation Forests for Symbolic Data as a Tool for Outlier Mining</b> . . . . .	150
Andrzej Dudek and Marcin Pelka	
<b>Symbolic Clustering Methods Applied to Interval Estimates of Production Cost Quantiles</b> . . . . .	151
Dominique Desbois	

<b>Fisher Discriminant Analysis for Interval Data</b> . . . . .	152
Diogo Pinheiro, Maria do Rosário Oliveira, Igor Kravchenko, and Lina Oliveira	
<b>Stability of Mixed-type Cluster Partitions for Determination of the Number of Clusters</b> . . . . .	153
Rabea Aschenbruck, Gero Szepannek, and Adalbert Wilhelm	
<b>Multinomial Multilevel Models with Discrete Random Effects: a Multivariate Clustering Tool</b> . . . . .	154
Chiara Masci, Francesca Ieva, and Anna Maria Paganoni	
<b>PD-clustering for Mixed Data Type</b> . . . . .	155
Francesco Palumbo and Cristina Tortora	
<b>Anomaly Detection-based Under-sampling for Imbalanced Classification Problems</b> . . . . .	156
You-Jin Park, Chun-Yang Peng, Rong Pan, and Douglas C. Montgomery	
<b>Continuous Adaptation to Distribution Drifts Through Continual Learning in Manufacturing</b> . . . . .	157
Henrique Siqueira and Onay Urfalioglu	
<b>Detecting Anomalies with TADGAN: a Case Study</b> . . . . .	158
Inês Oliveira e Silva, Carlos Soares, Arlete Rodrigues, and Pedro Bastardo	
<b>A Trivariate Geometric Classification of Decision Boundaries for Mixtures of Regressions</b> . . . . .	159
Filippo Antonazzo and Salvatore Ingrassia	
<b>Clustering Rainfall by Simulated Annealing for Histogram Symbolic Data</b>	160
Alejandro Chacón and Javier Trejos	
<b>Statistical Assessment of Youth Inclusion in the National Labour Markets</b>	161
Beata Bal-Domańska	
<b>Barriers to Industry Digitization in Poland from the Perspective of High and Medium-high Technology Sector Enterprises</b> . . . . .	162
Elżbieta Sobczak, Marcin Pelka, and Karolina Pokorska	
<b>Kernel Smoothing-based Probability Contours for Tumour Segmentation</b>	163
Wenhui Zhang and Surajit Ray	
<b>Parameter Estimation for Mixtures of Linear Mixed Models: the EM, CEM and SEM Algorithms</b> . . . . .	164
Luísa Novais and Susana Faria	



<b>Genomic Prediction Using Machine Learning Methods: Performance Comparison on Synthetic and Empirical Data</b> . . . . .	165
Vanda Lourenço, Joseph Ogutu, Rui Rodrigues, and Hans-Peter Piepho	
<b>Pooled Mean and Confidence Interval Estimation Combining Different Sets of Summary Statistics</b> . . . . .	166
Flora Ferreira, José Soares, Fernanda Sousa, Filipe Magalhães, Isabel Ribeiro, and Pedro Pacheco	
<b>Prediction of Diabetes via Bayesian Network Classifier from Exposure to Environmental Polluting Chemicals Data</b> . . . . .	167
Rosy Oh, Hong Kyu Lee, Youngmi Kim Pak, and Man-Suk Oh	
<b>Model Performance Metrics for Sample Selection Bias Correction by Pseudo Weighting</b> . . . . .	168
An-Chiao Liu, Ton de Waal, Katrijn Van Deun, and Sander Scholtus	
<b>Some Factors that Influence the Nature of Road Traffic Accidents</b> . . . . .	169
Paulo Infante, Gonçalo Jacinto, Anabela Afonso, Leonor Rego, Vitor Nogueira, Paulo Quaresma, José Saias, Daniel Santos, Pedro Nogueira, Marcelo Silva, Rosalina Pisco Costa, Patrícia Gois, and Paulo Rebelo Manuel	
<b>Comparing Variable Selection Methods for High-dimensional Compositional Data in a Discriminant Analysis Context</b> . . . . .	170
Pepus Daunis-i-Estadella, Glòria Mateu-Figueras, Viktorie Nesrstová, Karel Hron, and Josep A. Martín-Fernández	
<b>Cluster Analysis and Genetic Risk Score in Age-related Macular Degeneration - the Coimbra Eye Study</b> . . . . .	171
Rita Coimbra	
<b>Sensor System for Standardizing Articulation Patterns According to Korean Phonemes</b> . . . . .	172
Seong Tak Woo and Da Hee Oh	
<b>Categorical Data Visualization and the Cressie-Read Divergence Statistic</b>	173
Eric J. Beh and Rosaria Lombardo	
<b>Biplots Based on Latent Variable Models in the Analysis of Ecological Communities</b> . . . . .	174
Jenni Niku and Sara Taskinen	
<b>Biplot Representation of Partial Least Squares Regression for Binary Responses</b> . . . . .	175
Laura Vicente-Gonzalez and Jose Luis Vicente-Villardón	
<b>Generalized Spatio-temporal Regression with PDE Penalization</b> . . . . .	176
Eleonora Arnone, Elia Cunial, and Laura M. Sangalli	

<b>Impact Point Selection in Semiparametric Bifunctional Models</b> . . . . .	177
Silvia Novo, Germán Aneiros, and Philippe Vieu	
<b>Latent Function-on-scalar Regression Models for Observed Sequences of Correlated Binary Data: a Restricted Likelihood Approach</b> . . . . .	178
Fatemeh Asgari and Valeria Vitelli	
<b>pcTVI: Parallel MDP Solver Using a Decomposition into Independent Chains</b> . . . . .	179
Jaël Champagne Gareau, Éric Beaudry, and Vladimir Makarenkov	
<b>Classification of Viral Pneumonia Images via Multiple Instance Learning</b> 180	
Antonio Fuduli, Matteo Avolio, Eugenio Vocaturo, and Ester Zumpano	
<b>Nonlinear Approaches for Multiple Instance Learning</b> . . . . .	181
Annabella Astorino, Matteo Avolio, and Antonio Fuduli	
<b>An Online Minorization-Maximization Algorithm</b> . . . . .	182
Hien Duy Nguyen, Florence Forbes, Gersende Fort, and Olivier Cappé	
<b>Frugal Gaussian Clustering of Huge Ombalanced Datasets Through a Bin-marginal Approach</b> . . . . .	183
Filippo Antonazzo, Christophe Biernacki, and Christine Keribin	
<b>Reinforced EM Algorithm Through Clever Initialization for Clustering with Gaussian Mixture Models</b> . . . . .	184
Joshua Tobin, Chin Pang Ho, and Mimi Zhang	
<b>Exact Computation of the Angular Halfspace Depth</b> . . . . .	185
Stanislav Nagy and Rainer Dyckerhoff	
<b>Reconstruction of Atomic Measure Based on its Simplicial Depth</b> . . . . .	186
Petra Laketa and Stanislav Nagy	
<b>Robustness Aspects of Optimized Centroids</b> . . . . .	187
Jan Kalina and Patrik Janáček	
<b>Analysis of the Changes in the Polish Traditional Drugstores Market During COVID-19</b> . . . . .	188
Marcin Pelka, Antonio Irpino, and Michal Swachta	
<b>Logistic Regression Models for Aggregated Data</b> . . . . .	189
Thomas Whitaker, Boris Beranger, and Scott Sisson	
<b>Nonparametric Regressions for Distributional Data</b> . . . . .	190
Albert Meco, Javier Arroyo, and Antonio Irpino	

<b>Hotspot Cluster Detection Based on Spatial Hierarchical Structure and its Software</b> .....	191
Fumio Ishioka, Shoji Kajinishi, and Koji Kurihara	
<b>Group Lasso Penalty for Spatially Clustered Coefficient Regression</b> .....	192
Toshiki Sakai, Jun Tsuchida, and Hiroshi Yadohisa	
<b>Visualization of the Number of New Positives for COVID-19 in Japan</b> ...	193
Yoshiro Yamamoto, Sanetoshi Yamada, Mayumi Tanahashi, and Tadashi Imanishi	
<b>Two Simple but Efficient Algorithms to Recognize Robinson Dissimilarities</b> .....	194
Mikhaël Carmona, Guylain Naves, Victor Chepoi, and Pascal Pr�a	
<b>Clustering with Missing Data: Which Imputation Model for Which Cluster Analysis Method?</b> .....	195
Vincent Audigier, Nd�ye Niang, and Matthieu Resche-Rigon	
<b>Hierarchies and Weak-hierarchies as Interval Convexities</b> .....	196
Patrice Bertrand and Jean Diatta	
<b>A Rule-based Approach to Scoring Systems</b> .....	197
Michael Rapp, Johannes F�rnkranz, and Eyke H�llermeier	
<b>On Explaining Model Change Based on Feature Importance</b> .....	198
Maximilian Muschalik, Fabian Fumagalli, Barbara Hammer, and Eyke H�llermeier	
<b>Interpretable Multi-class Trees for Travel Choice Mode Analysis</b> .....	199
Christian Riccio, Andrea Papola, Michele Staiano, and Roberta Siciliano	
<b>Identification of Driver Genes in Glioblastoma via Regularized Classification</b> .....	200
Marta Belchior Lopes and Susana Vinga	
<b>Outlier Detection: a Procedure to Capture Atypical Groups of Observations</b> .....	201
Ana Helena Tavares, Vera Afreixo, and Paula Brito	
<b>Bayesian Classification and Non-Bayesian Label Estimation via EM Algorithm to Identify Differential Expression in Omics Data: a Comparative Study</b> .....	202
Marilia Antunes and Lisete Sousa	
<b>Sequence-aware Item Recommendations for Multiply Repeated User-item Interactions</b> .....	203
Juan Pablo Equihua, Maged Ali, Henrik Nordmark, and Berthold Lausen	

Contents	xxxi
<b>High-dimensional Linear Regression Estimation</b> . . . . .	204
Mauro Iannuzzi and Matteo Farnè	
<b>Experimental Study of Similarity Measures for Clustering Uncertain Time Series</b> . . . . .	205
Michael Dinzinger, Michael Franklin Mbouopda, and Engelbert Mephu Nguifo	
<b>Assessing the Status of Two Data-limited Skates Landed in Portuguese Ports Using an Empirical Catch Rule</b> . . . . .	206
Erick Chatalov, Ivone Figueiredo, Lisete Sousa, and Bárbara Pereira	
<b>Machine Learning Approach to Identify Factors that Influence Accident Severity</b> . . . . .	207
Daniel Santos, Vitor Nogueira, José Saias, Paulo Quaresma, Paulo Infante, Gonçalo Jacinto, Anabela Afonso, Leonor Rego, Pedro Nogueira, Marcelo Silva, Rosalina Pisco Costa, Patrícia Gois, and Paulo Rebelo Manuel	
<b>Trade and Bank Credit of Portuguese SMEs: a Panel Data Application</b> . .	208
Carla Henriques, Pedro Pinto, and Carolina Cardoso	
<b>Hausdorff Distance: a Powerful Tool for Matching Households and Individuals in Historical Censuses</b> . . . . .	209
Thais Pacheco Menezes, Michael Fop, and Thomas Brendan Murphy	
<b>Model-based Tri-clustering</b> . . . . .	210
Rafika Boutalbi, Lazhar Labiod, and Mohamed Nadif	
<b>Analyzing the Effects of Deviations from Normality on the Latent Growth Curve Models Goodness-of-fit</b> . . . . .	211
Catarina Marques, Maria de Fátima Salgueiro, and Paula C.R. Vicente	
<b>Transformation Mixture Modeling for Skewed Data Groups with Heavy Tails</b> . . . . .	212
Yana Melnykov, Xuwen Zhu, and Volodymyr Melnykov	
<b>A Simulation Study on Variable Selection in Mixture Regression Models</b> .	213
Susana Faria	
<b>Hybrid Forecasting Combinations by Feature Based Metalearning</b> . . . . .	214
Moises Santo, Andre C.P.L.F. de Carvalho, and Carlos Soares	
<b>On the Measurement of Household Subjective Poverty: Concepts and Application</b> . . . . .	215
Aleksandra Łuczak and Sławomir Kalinowski	

<b>Socio-economic Classification of Territorial Units: Extreme Value Theory-based Methods as Support for the Construction of a Synthetic Index</b> . . . . .	216
Aleksandra Łuczak and Małgorzata Just	
<b>Depth-based Two-sample Testing</b> . . . . .	217
Felix Gnettner, Claudia Kirch, and Alicia Nieto-Reyes	
<b>The Control of False Discovery Rate for Functional Data</b> . . . . .	218
Niels Lundtorp Olsen, Alessia Pini, and Simone Vantini	
<b>Functional Random Forest for Biomedical Signals Classification and Interpretative Tools</b> . . . . .	219
Fabrizio Mauro and Rosanna Verde	
<b>Correlation-based Iterative Clustering Methods for Time Course Data</b> . .	220
Michelle Carey, Shuang Wu, Guojun Gan, and Hulin Wu	
<b>Depth-based Classifiers for Partially Observed Functional Data</b> . . . . .	221
Antonio Elías, Raúl Jiménez, Anna Maria Paganoni, and Laura M. Sangalli	
<b>Using Clustering and Machine Learning Methods to Provide Intelligent Grocery Shopping Recommendations</b> . . . . .	222
Nail Chabane, Mohamed Achraf Bouaoune, Reda Amir Sofiane Tighilt, Bogdan Mazoure, Nadia Tahiri, and Vladimir Makarenkov	
<b>Typology of Motivation Factors for Employees in the Banking Sector: Multivariate Data Analysis</b> . . . . .	223
Áurea Sousa, Osvaldo Silva, M. Graça Batista, Sara Cabral, and Helena Bacelar-Nicolau	
<b>Industry Sector Detection in Legal Articles Using Transformer-based Deep Learning</b> . . . . .	224
Hui Yang, Stella Hadjiantoni, Yunfei Long, Rūta Petraitytė, and Berthold Lausen	
<b>User Segmentation Based on Online Behavioural Data via Ensemble Predictions and Clustering</b> . . . . .	225
Stella Hadjiantoni, Hui Yang, Yunfei Long, Ruta Petraityte, and Berthold Lausen	
<b>Attitudes Toward Statistics in the 3rd Cycle of Basic Education in Portugal</b>	226
Adelaide Freitas, Ana Julieta Morais, Pedro Sá Couto, and Anabela Rocha	
<b>Predictors of Quantitative Skills in Degree Schemes at University</b> . . . . .	227
Alex Partner, Adi Lausen, Alexei Vernitski, Chris Saker, and Berthold Lausen	

<b>Using Excel and R for Teaching Statistics and Data Analysis</b> . . . . .	228
W.H. Moolman	
<b>Students' Assessment Through a IRT and Archetypal Analysis Joint Strategy</b> . . . . .	229
Lucio Palazzo and Francesco Palumbo	
<b>Kernel-based Hierarchical Structural Component Models for Pathway Analysis</b> . . . . .	230
Suhyun Hwangbo, Sungyoung Lee, Seungyeoun Lee, Heungsun Hwang, Inyoung Kim, and Taesung Park	
<b>Bayesian Inference for the Generation Interval of COVID-19 in Busan, Korea</b> . . . . .	231
Jayoeng Paek, Iisu Choi, Kyeongah Nah, and Yongkuk Kim	
<b>Fitting an Accelerated Failure Time Model with Time-dependent Covariates via Nonparametric Gaussian Scale Mixtures</b> . . . . .	232
Ju-Young Park, Byungtae Seo, and Sangwook Kang	
<b>Comparison of Survival Prediction Models for Pancreatic Cancer: Cox Model vs. Machine Learning Models</b> . . . . .	233
Hyunsuk Kim, Taesung Park, and Seungyeoun Lee	
<b>Clustering High-dimensional Microbiome Data</b> . . . . .	234
Sanjeena Dang (Subedi) and Wangshu Tu	
<b>Clustering Adolescent Female Physical Activity Levels with an Infinite Mixture Model on Random Effects</b> . . . . .	235
Amy LaLonde, Tanzy Love, Deborah Rohm Young, and Tongtong Wu	
<b>Modeling Three-way RNA Sequencing Data Using Data Transformations and Matrix-variate Gaussian Mixture Models</b> . . . . .	236
Theresa Scharl and Bettina Grün	
<b>Some Issues in Robust Clustering</b> . . . . .	237
Christian Hennig	
<b>Assessing Common Principal Directions</b> . . . . .	238
David Rodríguez Vítóres and Carlos Matrán Bea	
<b>Robustness and Initialization Issues in Subspace Clustering</b> . . . . .	239
Luis A. García-Escudero and Agustín Mayo-Iscar	
<b>A Likelihood Ratio Test for Choosing Input Parameters in Robust Model Based Clustering</b> . . . . .	240
Luis A. García-Escudero, Agustín Mayo-Iscar, Gianluca Morelli, and Marco Riani	

<b>Data Clustering and Representation Learning Based on Networked Data</b>	241
Lazhar Labiod and Mohamed Nadif	
<b>Exploratory Graph Analysis for Configurational Invariance Assessment of a Test</b>	242
Alex Cucco, Lara Fontanella, Sara Fontanella, and Nicola Pronello	
<b>An Extension of Edge Reduction for Large Networks</b>	243
Pedro Campos	
<b>Patterns of Cooperation for Polish Authors of Research Publications in Economics, Business and Medicine Areas</b>	244
Urszula Cieraszewska, Paweł Lula, Magdalena Talaga, and Marcela Zembura	
<b>A Topological Clustering of Individuals</b>	245
Rafik Abdesselam	
<b>Modeling a Most Specific Generalization in Domain Taxonomies</b>	246
Zhirayr Hayrapetyan, Boris Mirkin, Susana Nascimento, Trevor Fenner, and Dmitry Froto	
<b>A Proposal for Formalization and Definition of Anomalies in Dynamical Systems</b>	247
Jan Michael Spoor, Jens Weber, and Jivka Ovtcharova	
<b>Unsupervised Classification of Categorical Time Series Through Innovative Distances</b>	248
Ángel López-Oriona, José A. Vilar, and Pierpaolo D'Urso	
<b>Detecting Differences in Italian Regional Health Services During Two Covid-19 Waves</b>	249
Lucio Palazzo and Riccardo Ievoli	
<b>The Clustering Performance of a Weighted Combined Distance Between Time Series</b>	250
Margarida G. M. S. Cardoso, Ana Alexandra Martins, and João Lagarto	
<b>Dimensionality Reduction and Multivariate Time Series Classification</b>	251
Veronne Yepmo, Angeline Plaud, and Engelbert Mephu Nguifo	
<b>A Review on Official Survey Item Classification for Mixed-Mode Effects Adjustment</b>	252
Afshin Ashofteh and Pedro Campos	
<b>Adaptive Fuzzy Systems in Economics and Finance: Evaluating Interval Forecasts of High-frequency Data</b>	253
Rosangela Ballini	

<b>The Usefulness of Selected Machine Learning Methods for Estimating Missing Data to Supplement Databases Used for Corporate Bankruptcy Prediction</b> . . . . .	254
Barbara Pawelek and Jozef Pocięcha	
<b>Registration of 24-hour Accelerometric Rest-activity Profiles and its Application to Human Chronotypes</b> . . . . .	255
Erin I. McDonnell, Vadim Zipunnikov, Jennifer A. Schrack, Jeff Goldsmith, and Julia Wrobel	
<b>Functional Data from Wearable Devices: a Review</b> . . . . .	256
Nihan Acar-Denizli and Pedro Delicado	
<b>A Wavelet-mixed Effect Landmark Model for the Effect of Potassium and Biomarkers Profiles on Survival in Heart Failure Patients</b> . . . . .	257
Caterina Gregorio, Giulia Barbatì, and Francesca Ieva	
<b>True Sparsity Approaches in Classification via Conic Optimization</b> . . . . .	258
Immanuel M. Bomze and Bo Peng	
<b>Creating Homogeneous Sectors: Criteria and Applications of Sectorization</b> . . . . .	259
Cristina Lopes, Maria Margarida Lima, Elif Göksu Öztürk, Ana Maria Rodrigues, Ana Catarina Nunes, Cristina Oliveira, José Soeiro Ferreira, and Pedro Filipe Rocha	
<b>MARGOT: a Maximum MARGin Optimal Classification Tree</b> . . . . .	260
Federico D'Onofrio, Marta Monaci, Giorgio Grani, and Laura Palagi	
<b>Multivariate Mapping of Soil Organic Carbon and Nitrogen</b> . . . . .	261
Stephan van der Westhuizen, David P. Hofmeyr, and Gerard B.M. Heuvelink	
<b>Spatial Configuration of Fire Stations in Portugal</b> . . . . .	262
Regina Bispo, Clara Yokochi, Francisca G. Vieira, Nádia Bachir, Pedro Espadinha-Cruz, José Pedro Lopes, Alexandre Penha, Marta Belchior Lopes, Filipe J. Marques, João Paulo Rodrigues, and António Grilo	
<b>Spatio-temporal Variability of Distribution and Abundance of Sardine in Portuguese Continental Coast: Environmental Effects</b> . . . . .	263
Daniela Silva, Raquel Menezes, Ana Moreno, Ana Teles-Machado, and Susana Garrido	
<b>Time Resolved Feature Importance of a Biopharmaceutical Purification Process Using Permutation Based Methods</b> . . . . .	264
Matthias Medl, Theresa Scharl, Astrid Dürauer, and Friedrich Leisch	



<b>Off-target Predictions in CRISPR-Cas9 Gene Editing Using Machine Learning</b> . . . . .	265
Ali Mertcan Kose and Ozan Kocadağlı	
<b>Comparison of k-mer and Alignment-based Pre-processing Approaches for Machine Learning Based Functional Annotation with 16S rRNA Data</b>	266
Rafal Kulakowski, Adi Lausen, Etienne Low-Decarie, and Berthold Lausen	
<b>An Ultrametric Model for Clustering and Dimensionality Reduction</b> . . . .	267
Giorgia Zaccaria	
<b>Combining Latent Class Analysis and Multiple Correspondence Analysis</b>	268
Alice Barth	
<b>Simultaneous Factorial Reduction and Clustering for Three-mode Data Sets: a Comparison</b> . . . . .	269
Prosper Ablordeppey, Adelaide Freitas, Maurizio Vichi, and Giorgia Zaccaria	
<b>Clustering Intensive Longitudinal Data Through Mixture Multilevel Vector-autoregressive Modeling</b> . . . . .	270
Anja Ernst, Marieke Timmerman, Feng Ji, Bertus Jeronimus, and Casper Albers	
<b>Mispecification Tests for Hidden Markov Models Based on a New Class of Finite Mixture Models</b> . . . . .	271
Francesco Bartolucci, Silvia Pandolfi, and Fulvia Pennoni	
<b>Natural Cubic Smoothing Splines for Latent Class Identification in Longitudinal Growth Trajectories</b> . . . . .	272
Katerina M. Marcoulides and Laura Trinchera	
<b>Supervised Classification via Neural Networks for Replicated Point Patterns</b> . . . . .	273
Kateřina Pawlasová, Iva Karafiátová, and Jiří Dvořák	
<b>Reliability Assessment of Ancient Stone Arch Bridge Applying ANN models, Case Study: Leça Railway Bridge</b> . . . . .	274
Edward A. Baron, Ana Margarida Bento, José Campos e Matos, Rui Calçada, and Kenneth Gavin	
<b>Application of Artificial Intelligence (AI) in Flood Risk Forecasting</b> . . . . .	275
Minh Quang Tran, Ana Margarida Bento, Elisabete Teixeira, Hélder Sousa, and José Campos e Matos	
<b>Logistic Regression with Sparse Common and Distinctive Covariates</b> . . .	276
Soogeun Park, Eva Ceulemans, and Katrijn Van Deun	

<b>Accuracy Measures for Binary Classification Based on Quantitative Group Tests</b> .....	277
Rui Santos, João Paulo Martins, and Miguel Felgueiras	
<b>Exploiting Pareto Density Estimation for Nonparametric Naïve Bayes Classifiers</b> .....	278
Quirin Stier and Michael C. Thrun	
<b>Author Index</b> .....	279

# On Parsimonious Modelling via Matrix-variate $t$ Mixtures

Salvatore D. Tomarchio

Mixture models for matrix-variate data have becoming more and more popular in the most recent years. One issue of these models is the potentially high number of parameters. To address this concern, parsimonious mixtures of matrix-variate normal distributions have been recently introduced in the literature. However, when data contains groups of observations with longer-than-normal tails or atypical observations, the use of the matrix-variate normal distribution for the mixture components may affect the fitting of the resulting model. Therefore, we consider a more robust approach based on the matrix-variate  $t$  distribution for modeling the mixture components. To introduce parsimony, we use the eigen-decomposition of the components scale matrices and we allow the degrees of freedom to be equal across groups. This produces a family of 196 parsimonious matrix-variate  $t$  mixture models. Parameter estimation is obtained by using an AECM algorithm. The use of our parsimonious models is illustrated via a real data application, where parsimonious matrix-variate normal mixtures are also fitted for comparison purposes.

**Keywords:** matrix-variate, mixture models, clustering, parsimonious models

---

Salvatore D. Tomarchio  
University of Catania, Department of Economics and Business, Catania, Italy  
e-mail: [daniele.tomarchio@unict.it](mailto:daniele.tomarchio@unict.it)