IFCS 2022

17th 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

Contents

Part V Invited Sessions

Monitoring Hyperparameter Choice for Robust Cluster Weighted Model

Andrea Cappozzo, Luis A. García-Escudero, Francesca Greselin, and Agustín Mayo-Iscar

The estimation of the Cluster Weighted Model is particularly attractive for providing explicit modeling of the explanatory variables, in a mixture of regression. The Robust version of the model requires the specification of a set of crucial parameters, like the proportion of trimmed units α, the thresholds to be adopted for the constrained estimation of groups scatter and for regression errors, beyond the number of components of the Mixture. To assist the choice of such hyper-parameters, a monitoring methodology could be of great help. The purpose is to provide a set of graphical tools to guide the final user in making an informed judgment, considering a landscape of plausible choices. The final output offers a set of optimal solutions, featured by the interval of hyper-parameters values in which their optimality holds, their stability and validity. An assessment of the role and extent of the outlying observations has been provided, introducing three new silhouette plots. The purpose is to understand the possible effects of the contaminated observations, with respect to the clustering of the covariate X, and the local regression lines Y, following the nature of the Cluster Weighted model.

Keywords: cluster-weighted modeling, outliers, trimmed bic, eigenvalue constraint, monitoring, model-based clustering, robust estimation

References

1. Cappozzo, A., García Escudero, L.A., Greselin, F., and Mayo-Iscar, A.: Parameter Choice, Stability and Validity for Robust Cluster Weighted Modeling. Stats **4** (3), 602–615 (2021)
2. Riani, M., Atkinson, A. C., Cerioli, A., and Corbellini, A.: Efficient robust methods via monitoring for clustering and multivariate data analysis. Pattern Recognit. **88**, 246?260, (2019)
3. Rousseeuw, P. J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. **20**, 53–65 (1987)

Andrea Cappozzo
Department of Mathematics, Politecnico di Milano, e-mail: andrea.cappozzo@polimi.it

Luis A. García Escudero
Departamento de Estadística e Investigación Operativa, Universidad de Valladolid,
e-mail: lagarcia@uva.es

Francesca Greselin
Department of Statistics and Quantitative Methods, University of Milano-Bicocca,
e-mail: francesca.greselin@unimib.it

Agustín Mayo-Iscar
Departamento de Estadística e Investigación Operativa, Universidad de Valladolid,
e-mail: agustin.mayo.iscar@uva.es

IFCS 2022

17th Conference of the International Federation of Classification Societies

Classification and Data Science in the Digital Age

Book of Abstracts