

SEMINAR

CEAUL & CEMAT



NEW GRAPHICAL DISPLAYS FOR CLASSIFICATION

ABSTRACT:

Classification is a major tool of statistics and machine learning. Several classifiers have interesting visualizations of their inner workings. Here we pursue a different goal, which is to visualize the cases being classified, either in training data or in test data. An important aspect is whether a case has been classified to its given class (label) or whether the classifier wants to assign it to a different class. This is reflected in the probability of the alternative class (PAC). A high PAC indicates label bias, i.e. the possibility that the case was mislabeled. The PAC is used to construct a silhouette plot which is similar in spirit to the silhouette plot for cluster analysis. The average silhouette width can be used to compare different classifications of the same dataset. We will also draw quasi residual plots of the PAC versus a data feature, which may lead to more insight in the data. One of these data features is how far each case lies from its given class, yielding so-called class maps. The proposed displays are constructed for discriminant analysis, k-nearest neighbors, support vector machines, CART, random forests, and neural networks. The graphical displays are illustrated and interpreted on data sets containing images, mixed features, and texts.

Joint work with: Jakob Raymaekers, Mia Hubert



14:30 AM



April 19th, 2023
Wednesday



IST, Maths. Building,
Room P3.10

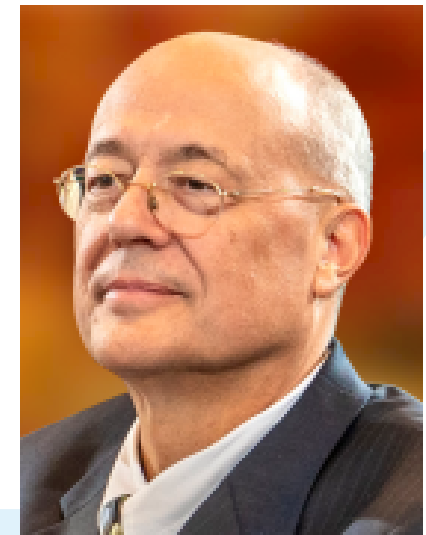
ZOOM

fct
Fundação
para a Ciência
e a Tecnologia

Project UIDB/00006/2020

CEAUL
Centro de Estatística e Aplicações
Universidade de Lisboa

ce mat
center for computational
and stochastic mathematics



SPEAKER

Peter Rousseeuw
KU Leuven, Belgium

Peter J. Rousseeuw scientific interests lie mostly in Statistics, Outlier, Estimator, Algorithm and Multivariate statistics. In his study, which falls under the umbrella issue of Statistics, Local regression is strongly linked to Econometrics. Peter J. Rousseeuw has researched Outlier in several fields, including Covariance and Principal component analysis, Pattern recognition.

Peter J. Rousseeuw works mostly in the field of Pattern recognition, limiting it down to concerns involving Cluster analysis and, occasionally, Theoretical computer science. His Estimator course of study focuses on Robustness and Data point. His studies in Robust regression integrate themes in fields like Least trimmed squares and Simple linear regression.