SCUOLA POLITECNICA E DELLE SCIENZE DI BASE
SCUOLA DELLE SCIENZE UMANE E SOCIALI

PHD PROGRAM IN INDUSTRIAL ENGINEERING
CURRICULUM ENGINEERING MANAGEMENT, QUALITY AND DATA SCIENCE FOR TECHNOLOGY

ANOMALY DETECTION BY ROBUST STATISTICS
Prof. Mia Hubert, Prof. Peter J. Rousseeuw
KU Leuven, Belgium
Section of Statistics and Data Science, Department of Mathematics

AIM OF THE COURSE: The goal of robust statistics is to develop and study techniques for data analysis that are resistant to outlying measurements, and are also able to detect these anomalies.
In this course, we introduce notions of robustness such as the breakdown value and the sensitivity function. We study several robust estimators of univariate location and scale, multivariate location and covariance, linear regression, and principal component analysis as well as robust classifiers. Special attention is given to diagnostic tools to flag anomalies.

LESSON CALENDAR

- **Tuesday, 21 March 2023, 14:00-16:00 (Prof. Mia Hubert)**
  Room: Biblioteca Storica, Scuola Politecnica e delle Scienze di Base, II piano, P.le V. Tecchio 80 - Napoli
  Robust location and scale

- **Tuesday, 21 March 2023, 16:30-18:30 (Prof. Peter Rousseeuw)**
  Room: Biblioteca Storica, Scuola Politecnica e delle Scienze di Base, II piano, P.le V. Tecchio 80 -Napoli
  Robust covariance estimation and regression

- **Wednesday, 22 March 2023, 10:30-12:30 (Prof. Mia Hubert)**
  Room: Aula Convegni, Palazzo degli Uffici, I piano, Via Giulio Cesare Cortese, 29 - Napoli
  Robust principal component analysis (PCA) and classification

DETAILS: The course must be attended in person and will be held in the English language.

HOW TO ENROL IN THE COURSE: Participants must enrol the course by filling in the form available at the following link until the maximum capacity of the rooms is reached https://forms.office.com/e/g8Uy9dksLu

Mia Hubert is a professor at the KU Leuven, Section of Statistics and Data Science in the Mathematics Department. She is teaching several bachelor's and master's courses for students in science and engineering science. Her research focuses on robust statistics, outlier detection, data visualization, depth functions, and the development of statistical software. She is an elected fellow of the ISI and has served as an associate editor for several journals such as JCGS, CSDA, and Technometrics. She is co-founder and co-organizer of the new bi-annual Rousseeuw Prize for Statistics.
More information can be found at
https://en.wikipedia.org/wiki/Mia_Hubert
https://scholar.google.com/citations?user=ZgTLgMYAAAAJ&hl=en
http://wis.kuleuven.be/statdatascience/robust

Peter Rousseeuw is a Belgian statistician who has mainly contributed to robust methodology, outlier detection, and cluster analysis. He obtained his PhD following research at ETH Zurich. Later he was a professor at the Delft University of Technology and at the University of Antwerp, Belgium. Afterward, he spent over a decade in the financial industry in New York. Currently, he is professor emeritus at KU Leuven, Belgium. Among the techniques he introduced are the Least Trimmed Squares method for robust regression and the Minimum Covariance Determinant method for covariance matrices. With Leon Kaufman he coined the word medoid when proposing the k-medoids method for cluster analysis. His silhouette display for clustering is often used to select the number of clusters. Other works are on robust principal component methods, regression depth, the bagplot (with John Tukey), and dealing with cellwise outliers. His work has been cited over 100,000 times. More information can be found at
https://en.wikipedia.org/wiki/Peter_Rousseeuw
https://scholar.google.com/citations?user=5LMM6rsAAAAJ&hl=en

SCUOLA POLITECNICA E DELLE SCIENZE DI BASE
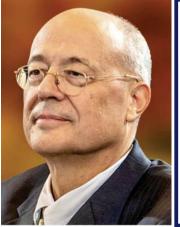SCUOLA DELLE SCIENZE UMANE E SOCIALI

Prof. Peter J. Rousseeuw

KU Leuven, Belgium
Section of Statistics and Data Science, Department of Mathematics

Visualizing Classification Results

(joint work with Jakob Raymaekers and Mia Hubert)

Wednesday, March 22nd, 2023, 14:30 - 15:30
Università degli Studi di Napoli Federico II
Aula Convegni, Palazzo degli Uffici, I Piano
Via Giulio Cesare Cortese, 29 - Napoli

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.

Peter Rousseeuw is a Belgian statistician who has mainly contributed to robust methodology, outlier detection, and cluster analysis. He obtained his PhD following research at ETH Zurich. Later he was professor at the Delft University of Technology and at the University of Antwerp, Belgium. Afterward he spent over a decade in the financial industry in New York. Currently he is professor emeritus at KU Leuven, Belgium. Among the techniques he introduced are the Least Trimmed Squares method for robust regression and the Minimum Covariance Determinant method for covariance matrices. With Leon Kaufman he coined the word medoid when proposing the k-medoids method for cluster analysis. His silhouette display for clustering is often used to select the number of clusters. Other works are on robust principal component methods, regression depth, the bagplot (with John Tukey), and dealing with cellwise outliers. His work has been cited over 100,000 times.

More information can be found at
Wikipedia https://en.wikipedia.org/wiki/Peter_Rousseeuw
Google Scholar https://scholar.google.com/citations?user=5LMM6rsAAAAJ&hl=en

Participants must enrol the course by filling in the form available at the following link until the maximum capacity of the rooms is reached https://forms.office.com/e/g8Uy9dksLu

SCUOLA POLITECNICA E DELLE SCIENZE DI BASE
SCUOLA DELLE SCIENZE UMANE E SOCIALI

Prof. Mia Hubert

KU Leuven, Belgium
Section of Statistics and Data Science, Department of Mathematics

Cellwise robust methods

(joint work with Peter Rousseeuw, Wannes Van den Bossche and Mehdi Hirari)

Thursday, March 23rd, 2023, 10:30 - 11:30,
Università degli Studi di Napoli Federico II
Aula Convegni, Palazzo degli Uffici, I Piano
Via Giulio Cesare Cortese, 29 - Napoli

Abstract – Multivariate data are typically represented by a rectangular matrix in which the rows are the objects and the columns are the variables. Much research has focused on handling rowwise outliers, i.e. rows that deviate from the majority of the rows in the data (for instance, they might belong to a different population). In recent years also cellwise outliers are receiving attention. These are suspicious cells that can occur anywhere in the data matrix. Even a relatively small proportion of outlying cells can contaminate over half the rows, which causes rowwise robust methods to break down. Several recent methods are presented that can handle rowwise and cellwise outliers simultaneously, as well as missing values. They are applied to multivariate data, high-dimensional data, functional data and multi-way data.

Mia Hubert is professor at the KU Leuven, Section of Statistics and Data Science in the Mathematics Department. She is teaching several bachelor and master courses for students in science and engineering science. Her research focuses on robust statistics, outlier detection, data visualization, depth functions, and the development of statistical software. She is an elected fellow of the ISI and has served as associate editor for several journals such as JCGS, CSDA, and Technometrics. She is co-founder and co-organizer of the new bi-annual Rousseeuw Prize for Statistics.

More information can be found at
Wikipedia https://en.wikipedia.org/wiki/Mia_Hubert
Google Scholar https://scholar.google.com/citations?user=ZgTLgMYAAAAJ&hl=en
CV can be downloaded from http://wis.kuleuven.be/statdatascience/robust

Participants must enrol the course by filling in the form available at the following link until the maximum capacity of the rooms is reached https://forms.office.com/e/g8Uy9dksLu