DaSSWeb
Data Science and Statistics Webinar

24TH SEPTEMBER | 2.30PM (WEST) | JOIN US HERE

ROBUST PRINCIPAL COMPONENTS BY CASEWISE AND CELLWISE WEIGHTING

Principal component analysis (PCA) is a fundamental tool for analyzing multivariate data. Here the focus is on dimension reduction to the principal subspace, characterized by its projection matrix. The classical principal subspace can be strongly affected by the presence of outliers. Traditional robust approaches consider casewise outliers, that is, cases generated by an unspecified outlier distribution that differs from that of the clean cases. But there may also be cellwise outliers, which are suspicious entries that can occur anywhere in the data matrix. Another common issue is that some cells may be missing. This paper proposes a new robust PCA method, called cellPCA, that can simultaneously deal with casewise outliers, cellwise outliers, and missing cells. Its single objective function combines two robust loss functions that together mitigate the effect of casewise and cellwise outliers. The objective function is minimized by an iteratively reweighted least squares (IRLS) algorithm. Residual cellmaps and enhanced outlier maps are proposed for outlier detection. Simulations and real data examples illustrate the performance of cellPCA.
Authors: Fabio Centofanti, Mia Hubert, Peter Rousseeuw

SPEAKER

PETER J. ROUSSEEUW | DEPARTMENT OF MATHEMATICS, KU LEUVEN, LEUVEN, BELGIUM

Peter Rousseeuw is known mainly for his work on robust statistics. Among his creations are least trimmed squares regression, the minimum covariance determinant estimator, the k-medoids clustering method, and the silhouettes graphical display. Peter obtained his PhD in 1981 following research carried out at the ETH in Zürich, Switzerland, which led to a book on influence functions. Later, he was a professor at Delft University of Technology, The Netherlands, and at the University of Antwerp, Belgium. Next, he was a researcher at Renaissance Technologies in New York for over a decade. He then returned to Belgium as a professor at KU Leuven, until becoming emeritus in 2022. He is an elected member of ISI and a fellow of IMS and ASA. In the course of his career, Peter published three books and over 200 papers on theory, algorithms and applications, together receiving over 115,000 citations. He was awarded the George Box Medal for Business and Industrial Statistics, the Research Medal of the International Federation of Classification Societies, the Frank Wilcoxon Prize, and twice the Jack Youden Prize. Recently, Peter received the 2024 ASA Noether Distinguished Scholar Award for nonparametric statistics. His former PhD students include Annick Leroy, Rik Lopuhaa, Geert Molenberghs, Christophe Croux, Mia Hubert, Stefan Van Aelst, Tim Verdonck and Jakob Raymaekers. Peter's recent work is mainly on robustness to cellwise outliers.

FEP