dmat universidade de aveiro departamento de matemática

fct Fundação para a Ciência e a Tecnologia

Seminário

Grupo de Probabilidades e Estatística

7 de outubro de 2024 15:00 Sala Sousa Pinto

Outlier detection and explanation for matrix-variate data

Peter Filzmoser[1], Marcus Mayrhofer[1], Una Radojicic[1], Horst Lewitschnig[2]

[1]: *TU Wien – Vienna University of Technology, Austria*

[2]: *Infineon Technologies, Austria AG*

Abstract

In multivariate data analysis we consider an observation as a vector, with measurements for each variable. Outlier detection techniques are well established for this case. Here we extend vector-valued observations to matrix-valued objects, and thus the measurements of an observation are arranged in the rows and columns of a matrix. A prominent example for such a data structure are image data, where the pixel information is typically presented in a rectangular matrix. In traditional analyses, the pixel information is converted row-by-row (or column-by-column) to a long vector, which is considered as a row in a high-dimensional matrix of observations. With such a representation, the information about the neighborhood relationships of the pixels is lost.

The concept of matrix-valued data is not new at all, and a prominent distribution in this context is the matrix normal distribution. There are different proposals in the literature on how to estimate the parameters of this distribution. It is also possible to define a Mahalanobis distance, and the concept of robust covariance estimation can be modified to obtain robust estimators for the matrix-valued case. We present an adaptation of the well-known MCD (Minimum Covariance Determinant) estimator to this situation. Moreover, the concept of Shapley values, which has been successfully used in the context of Explainable AI, is extended in order to explain the reasoning behind the outlyingness. In the context of image data, for example, one can identify outlying images and explain which pixels contribute to this outlyingness.

A more detailed background, as well as illustrative examples, will be provided in the presentation.

This seminar is supported by CIDMA – Center for Research and Development in Mathematics and Applications through FCT – Fundação para a Ciência e a Tecnologia, within projects UIDB/04106/2020 and UIDP/04106/2020, with links https://doi.org/10.54499/UIDB/04106/2020 and https://doi.org/10.54499/UIDP/04106/2020.

CIDMA CENTRO DE I&D EM MATEMÁTICA E APLICAÇÕES CENTER FOR R&D IN MATHEMATICS AND APPLICATIONS

REPÚBLICA PORTUGUESA