

Handling Cellwise Outliers by Sparse Regression and Robust

Covariance

Jakob Raymaekers

KU Leuven

Abstract

We propose a data-analytic method for detecting cellwise outliers. Given a robust covariance matrix, outlying cells (entries) in a row are found by the cellHandler technique which combines lasso regression with a stepwise application of constructed cutoff values. The penalty term of the lasso has a physical interpretation as the total distance that suspicious cells need to move in order to bring their row into the fold. For estimating a cellwise robust covariance matrix, we construct a detection-imputation method which alternates between flagging outlying cells and updating the covariance matrix as in the EM algorithm. The proposed methods are illustrated by simulations and on real data about volatile organic compounds in children.

Keyword: Discriminant analysis; k-Nearest neighbors; Mislabeling; Pairwise coupling; Support vector machines