Abstract

In this thesis, challenges inherent to analyzing datasets from empirical experiments are addressed, with a particular focus on outlier detection and the analysis of high-dimensional data. By proposing robust approaches and leveraging modern optimization algorithms, the thesis offers practical solutions for enhancing the reliability and efficiency of data analysis techniques in complex real-world scenarios, offering valuable insights and practical tools for researchers and practitioners in various fields. Contributions to robust and sparse regression, association, and dimension reduction are illustrated on datasets from tribology, a multi-disciplinary field studying friction, wear, and lubrication. These data result from practical experiments with engine oils in different conditions and from several degradation pathways and include spectral, functional, and image data with only a limited number of observations. Robust methods tailored for low-dimensional data do not suffice for handling experimental datasets in high-dimensional settings. Therefore, this thesis presents suitable preprocessing and sampling strategies for robust regression and classification in this challenging setting. In addition, a combination of robust statistical methods with gradient-based optimization techniques is proposed for quantifying the relation between two multivariate datasets using robust and sparse CCA (canonical correlation analysis) and dimension reduction via robust and sparse PCA (principal component analysis).