

Robust sparse regression in high dimensions

Andreas Alfons

Erasmus University Rotterdam, Rotterdam, The Netherlands alfons@ese.eur.nl,
KU Leuven, Leuven, Belgium

Christophe Croux

KU Leuven, Leuven, Belgium christophe.croux@kuleuven.be

Viktoria Öllerer

KU Leuven, Leuven, Belgium viktoria.oellerer@kuleuven.be

Sarah Gelper

Erasmus University Rotterdam, Rotterdam, The Netherlands sgelper@rsm.nl

Due to the increasing availability of data sets with a large number of variables, sparse model estimation is a topic of high importance in modern data analysis. Sparse regression allows to improve prediction performance by variance reduction and increase interpretability of the resulting models due to the smaller number of explanatory variables. Sparsity in the coefficient estimates can thereby be achieved through appropriate regularization techniques. Another common problem in applied statistics is the presence of outliers in the data. Adding a suitable penalty function on the coefficient estimates to the objective functions of robust regression estimators yields new estimators that are both sparse and robust. One example for this class of estimators is the sparse least trimmed squares estimator, or sparse LTS for short, which combines the least trimmed squares objective function with the lasso penalty. We compare the sparse LTS estimator with other sparse regression methods in numerical experiments. In addition, we derive influence functions for a general class of regularized regression estimators.

Key Words: Influence function, outliers, regularization, sparse least trimmed squares