

Multivariate Outlier Detection and Cluster Identification

David M. Rocke¹, and David L. Woodruff²

¹ Department of Applied Science, University of California, Davis, Davis, CA 95616

² Graduate School of Management, University of California, Davis, Davis, CA 95616

Keywords: Heuristic Search; Model-Based Clustering; Robust Multivariate Estimation

Abstract

We examine relationships between the problem of robust estimation of multivariate location and shape and the problem of maximum likelihood assignment of multivariate data to clusters. Recognition of the connections between estimators for clusters and outliers immediately yields one important result that we demonstrate in this paper; namely, outlier detection procedures can be improved by combining them with cluster identification techniques. Using this combined approach, one can achieve practical breakdown values that approach the theoretical limits. We report computational results that demonstrate the effectiveness of this approach. In addition, we provide a new robust clustering method.

Please fill in this form and mail it together with your abstract.

My abstract fits best to topic number ... 14 (**“PLEASE INDICATE HERE THE NUMBER FROM THE LIST OF TOPICS BELOW WHICH BEST FITS TO YOUR ABSTRACT”**).

List of Topics:

1. Algorithms
2. Applications
3. Biostatistics
4. Computing and graphics
5. Data analysis
6. Data mining
7. Economics, finance
8. Efficiency and robustness
9. Functionals and bias
10. Fuzzy statistics
11. Geostatistics
12. Inference for robust methods, model testing
13. Location depth and regression depth
14. Multivariate methods
15. Neural networks
16. Rank-based methods
17. Regression quantiles, trimming
18. Robust covariance
19. Robust designs
20. Robust regression
21. Time series analysis
22. Wavelets
23. Other (please specify)