Robust PCA with bootstrap based on MM-estimators

G. Willems[1], S. Van Aelst[2], and M. Salibián-Barrera[3]

[1] Department of Mathematics and Computer Science, University of Antwerp, Middelheimlaan 1, B-2020 Antwerp, Belgium

[2] Department of Applied Mathematics and Computer Science, Ghent University, Krijgslaan 281 S9, B-9000 Ghent, Belgium

[3] Department of Statistics, University of British Columbia, 6356 Agriculture Road, Vancouver, B.C., Canada V6T 1Z2

Keywords: Principal components analysis, Robust Bootstrap, MM-estimators.

Abstract

We consider robust principal components analysis (PCA) based on multivariate MM-estimators as introduced by Tatsuoka and Tyler (2000). The MM-estimators are designed to be both highly robust against outliers and highly efficient in case of normal data. Essentially, an S-estimator is used to obtain a robust scale estimate, after which the location and shape are estimated with a more efficient M-estimator.

We primarily focus on inference procedures based on bootstrap, as an alternative to inference based on asymptotic normality. Classical bootstrap for the MM-estimator has some major drawbacks: it can be extremely time-consuming and it has a robustness problem. We therefore consider the fast and robust bootstrap proposed by Salibian-Barrera and Zamar (2002). The bootstrap can be used to estimate the variability of eigenvalues, eigenvectors and other statistics of interest in PCA. Furthermore, it can be helpful in deciding the number of principal components to retain.

We present consistency results for the bootstrap and show its finite-sample accuracy as investigated through simulations. We also illustrate the use of the robust principal components method and the bootstrap inference on real data.

References

M. Salibián-Barrera and R. Zamar (2002). Bootstrapping robust estimates of regression. *The Annals of Statistics*, 30, 556–582.

K.S. Tatsuoka and D.E. Tyler (2000). The uniqueness of S and M-functionals under non-elliptical distributions. *The Annals of Statistics*, 28, 1219–1243.