Robust Principal Component Analysis: a Review

Subhabrata Majumdar, Snigdhansu Chatterjee

Abstract:

Keywords: Data depth; Multivariate ranking; Principal components analysis; robust statistics; functional data; dimension reduction

1 Introduction

Principal component Analysis (PCA) is one of the oldest, yet most widely used methods of unsupervised multivariate analysis. For a data matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$ containing observations in p variables for n samples, each column having mean 0, principal components are defined as p-dimensional vectors $\mathbf{w}_k, 1 \leq k \leq p$ such that

$$\mathbf{w}_k = \begin{cases} \arg\max_{\|\mathbf{w}\|=1} \mathbf{w}^T \mathbf{X}^T \mathbf{X} \mathbf{w} \text{ if } k = 1 \\ \arg\max_{\|\mathbf{w}\|=1} \mathbf{w}^T \mathbf{R}_k^T \mathbf{R}_k \mathbf{w} \text{ if } k > 1; \quad \mathbf{R}_k = \mathbf{X} - \sum_{s=1}^{k-1} \mathbf{X} \mathbf{w}_s \mathbf{w}_s^T \end{cases} \tag{1.1}$$

Following a lagrange multiplier approach, the eigenvectors of $\mathbf{X}^T\mathbf{X}$, equivalently the right singular vectors obtained from the singular value decomposition of $\mathbf{X}$ provide solutions to (1.1).

References

K. P. Adragni and R. D. Cook. Sufficient dimension reduction and prediction in regression. *Phil. Trans. R. Soc. A*, 367:4385–4405, 2009.

N. Locantore, J.S. Marron, D.G. Simpson, N. Tripoli, J.T. Zhang, and K.L. Cohen. Robust principal components of functional data. *TEST*, 8:1–73, 1999.

Y. Zuo and R. Serfling. General notions of statistical depth functions. *Ann. Statist.*, 28-2:461–482, 2000.