

Robust Principal Components for data with elementwise contamination

Ricardo Maronna and Víctor J. Yohai

Abstract

Consider fitting p principal components (PC) to a m -dimensional dataset given by the $n \times m$ matrix \mathbf{X} , where each row \mathbf{x}_i , $1 \leq i \leq n$, is a m -dimensional observation. Suppose that in addition to the possibility that some observations being atypical, each element x_{ij} has a probability δ of being affected by a gross error. This occurs frequently with high-dimensional data, in particular images and spectra. If $m\delta$ is large (say, >0.7), then there is a high probability that the number of rows \mathbf{x}_i containing at least a gross error be larger than $n/2$, which would cause the usual robust PC estimators to break down.

We consider two approaches to this problem, both based on a robust M-scale S . For a $n \times m$ matrix $\mathbf{Y} = [\mathbf{y}^1, \mathbf{y}^2, \dots, \mathbf{y}^m]$ put $V(\mathbf{Y}) = \sum_{j=1}^m S(\mathbf{y}^j)^2$.

The first approach is a robust approximation of \mathbf{X} by a matrix $\hat{\mathbf{X}}$ of rank p : $\hat{\mathbf{X}} = \mathbf{A}\mathbf{B}'$, where \mathbf{A} and \mathbf{B} are respectively $n \times p$ - and $m \times p$ -matrices, such that $V(\mathbf{X} - \hat{\mathbf{X}}) = \min$. This definition takes into account the possibility that the variability not explained by the principal components may differ among columns.

The second approach is a combination of S-estimation with replacement of suspicious observations by cleaned values. Let \mathbf{P} be the projection matrix on a p -dimensional subspace and let ψ be an odd bounded increasing function such that $\psi(u) = u$ if $|u| \leq c$ for some c . Then, define the $n \times m$ clean matrix $\tilde{\mathbf{X}} = (\tilde{x}_{ij})$, the $n \times m$ matrix $\hat{\mathbf{X}}$, with rows $\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_n$ and σ_j , $1 \leq j \leq m$, by $\tilde{x}_{ij} = \hat{x}_{ij} + \sigma_j \psi((x_{ij} - \hat{x}_{ij})/\sigma_j)$, $\hat{\mathbf{x}}_i = \mathbf{P}\tilde{\mathbf{x}}_i$ and σ_j is the scale S applied to the j -th column of $\mathbf{X} - \hat{\mathbf{X}}$. Then, \mathbf{P} is chosen such that $S(\|\mathbf{x}_1 - \hat{\mathbf{x}}_1\|, \dots, \|\mathbf{x}_n - \hat{\mathbf{x}}_n\|) = \min$.

In both cases we must approximate the global minimum of a highly nonlinear function. We propose algorithms based on a combination of random and gradient searches, which are adequately fast even for high m . Preliminary simulations show that both approaches are able to deal with highly contaminated situations in which the usual robust PC estimators fail.