ROBUST PRINCIPAL COMPONENT ANALYSIS USING ALPHA DIVERGENCE

ICIP 2020

Aref Miri Rekavandi, Abd-Krim Seghouane

The university of Melbourne

October, 2020

Outline

Model

Observations follow a general linear model (GLM) as

$$\mathbf{x} = \mathbf{V}\theta + \mathbf{n}$$

where

- $\mathbf{x} \in \mathbb{R}^m$
- $\mathbf{V} \in \mathbb{R}^{m \times r}$ [Orthonormal matrix]
- $\theta \in \mathbb{R}^r$ [unknown and $\sim \mathcal{N}(\mathbf{0}, \mathbf{I})$]
- $\mathbf{n} \in \mathbb{R}^m$, and nominally$\sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$

N samples in total are available and all samples are put in matrix $\mathbf{X} = \{\mathbf{x}_1, ..., \mathbf{x}_N\}$.

Problem Definition

The objective is to find r principal loading vectors such that they can explain most of the data variance.

$$\hat{\mathbf{u}}_k = \arg\max_{\mathbf{u}} \mathbf{Var}(\mathbf{u}^\top \mathbf{X}), \text{ s.t.} \mathbf{U}^\top \mathbf{U} = \mathbf{I}, \ k = 1, \cdots, r,$$

Or equivalently, using projection framework

$$\hat{\mathbf{U}} = \arg\min_{\mathbf{U}} \|\mathbf{X} - \mathbf{U}\mathbf{U}^\top \mathbf{X}\|_F \textbf{ s.t. } \mathbf{U}^\top \mathbf{U} = \mathbf{I}.$$

The solution of above problems is the first r eigenvectors of the sample covariance matrix $\mathbf{X}\mathbf{X}^\top$.
However, this solution is not robust against noise deviation!

Robust PCA

The PCA problem is considered as a rank minimization problem using the model[1]

$$\mathbf{X} = \mathbf{L} + \mathbf{S} + \epsilon,$$

where $\mathbf{L}$ and $\mathbf{S}$ are low rank and sparse matrices, respectively. Therefore, they are given by

$$\hat{\mathbf{L}}, \hat{\mathbf{S}} = \arg\min_{\mathbf{L},\mathbf{S}} \|\mathbf{L}\|_* + \lambda_1 \|\mathbf{S}\|_1 + \|\mathbf{X} - \mathbf{L} - \mathbf{S}\|_F.$$

where $\|\mathbf{L}\|_*$ denotes the nuclear norm of $\mathbf{L}$ which measures the sum of singular values of $\mathbf{L}$.
Still sensitive due to using non robust norms.

1. Candes et al., 2011: Robust principal component analysis?

From KL divergece to $\alpha-$divergence

Using the probabilistic PCA point of view, the solution of classical PCA is the maximum likelihood solution when a generative linear model and multivariate Gaussian noise is used. It is well known that when $(N \to \infty)$

$$\hat{\beta}_{ML} = \arg\max_{\beta} \frac{1}{N} \sum_{i=1}^{N} \log\left(f(\mathbf{x}_i, \beta)\right) = \arg\min_{\beta} KL\left(f(\mathbf{x}, \beta^*), f(\mathbf{x}, \beta)\right).$$

As an alternative we propose to use the $\alpha-$divergence defined as

$$D_\alpha\left(g(\mathbf{y}, \lambda) \parallel f(\mathbf{y}, \omega)\right) = \frac{1}{\alpha(\alpha-1)}\left[\int g(\mathbf{y}, \lambda)^\alpha f(\mathbf{y}, \omega)^{1-\alpha} d\mathbf{y} - 1\right],$$

to develop the estimator. When $\alpha \to 1$, $D_\alpha(.) \to KL(.)$.

Proposed Robust PCA

In order to estimate the mean and the covariance of data we have

$$\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}} = \arg\min_{\boldsymbol{\mu}, \boldsymbol{\Sigma}} D_\alpha \left(g_e(\mathbf{x}) \parallel \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \right),$$

where $g_e(\mathbf{x}) = \frac{1}{N} \sum_{n=1}^{N} \delta(\mathbf{x} - \mathbf{x}_n)$ is the empirical density and

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{m/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\},$$

Now with taking derivatives with respect to the mean and the covariance and putting them equal to zero we have:

$$\boldsymbol{\mu}^t = \frac{\sum_{n=1}^{N} w_n^t \mathbf{x}_n}{\sum_{n=1}^{N} w_n^t},$$

Proposed Robust PCA

$$\boldsymbol{\Sigma}^t = \frac{\sum_{n=1}^{N} w_n^t (\mathbf{x}_n - \boldsymbol{\mu}^{t-1})(\mathbf{x}_n - \boldsymbol{\mu}^{t-1})^\top}{\sum_{n=1}^{N} w_n^t}.$$

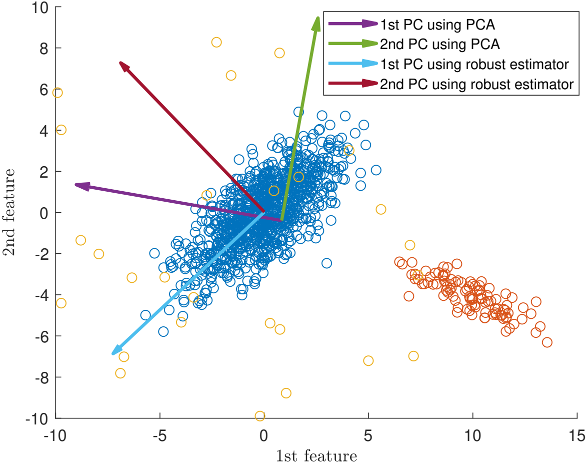
w_n^t shows the weight of $\mathbf{x}_n$ at step t in estimation of parameters and can be computed by $w_n^t = \exp\{-\frac{1-\alpha}{2}(\mathbf{x}_n - \boldsymbol{\mu}^{t-1})^\top (\boldsymbol{\Sigma}^{\mathbf{t-1}})^{-1}(\mathbf{x}_n - \boldsymbol{\mu}^{t-1})\}$.

Remarks:

- If data is not zero mean, first we should make it zero mean.
- The method is tunable by α and for $\alpha < 1$ the robustness property can be achieved.
- The method is an iterative algorithm which converges fast and depends on initial estimate.
- Number of observations (N) is more than the data dimension (m).

Results

Scatter plot of simulated data (blue, orange and yellow circles show the original data, structured and unstructured outliers, respectively) and estimated principal loading vectors using classical and robust PCA algorithm.



Results

Background-Foreground (FB) separation task on airport data (top row is the background and bottom row is the foreground). First 200 frames with the size of 72×88 is used. Results on frame number 8:

From Left: Original frame, PCA, GoDec, Proposed.