Parsimonious Feature Extraction Methods: Extending Robust Probabilistic Projections with Generalized Skew-t

Dorota Toczydlowska
School of Mathematical and Physical Sciences
University of Technology Sydney
dtoczydlowska@gmail.com

Gareth W. Peters
Department of Actuarial Mathematics and Statistics
Heriot-Watt University
garethpeters78@gmail.com

Pavel V. Shevchenko
Department of Actuarial Studies and Business Analytics
Macquarie University
pavel.shevchenko@mq.edu.au

October 12, 2020

Abstract

The study focuses on extension to the approach of Principal Component Analysis (PCA), as defined in [1]. PCA and related matrix factorisation methodologies are widely used in data-rich environments for dimensionality reduction, data compression, feature-extraction techniques or data de-noising. The methodologies identify a lower-dimensional linear subspace to represent the data, which captures second-order dominant information contained in high-dimensional data sets. PCA can be viewed as a matrix factorisation problem which aims to learn the lower-dimensional representation of the data, preserving its Euclidean structure. However, in the presence of either a non-Gaussian distribution of the data generating distribution or in the presence of outliers which corrupt the data, the standard PCA methodology provides biased information about the lower-rank representation.

In many applications, the stochastic noise or observation errors in the data set are assumed to be, in some sense, "well-behaved"; for instance, additive, light-tailed, symmetric and zero-mean. When non-robust feature extraction methods are naively utilised in the presence of violations of these implicit statistical assumptions, the information contained in the extracted features cannot be relied upon, resulting in misleading inference. Therefore, it is critical to ensure that the feature extraction captures information about correct characteristics of the process generating the data. In the following study, we relax the inherent assumption of "well-behaved" observation noise by developing a class of robust estimators that can withstand violations of such assumptions, which routinely arise in real data sets.

Many improvements to the classical PCA methodology have been introduced in the literature to accommodate various characteristics that may deviate from the standard assumptions on the data when applying classical PCA methods. For instance robust variants of standard PCA modify the distance measure between each observation and its lower-rank approximation. Majority of robust feature-extraction methods in the literature are primarily based on the assumption that observations are independent over time and that the marginal distributions of their components have the same profile of heavy tails. This reasoning might be criticised for having a limited ability to capture various tail-dependence patterns in multivariate data analysis. Therefore, we want to investigate an approach that will be able to accommodate a broader range of dependence and marginal distribution assumptions in the data-generating mechanism. To achieve that, we employ Probabilistic PCA (PPCA) which model has been introduced by [2]. The method seeks k - dimensional uncorrelated latent vector $\mathbf{X}_t$ which provides the most meaningful model of d-dimensional observed random vector $\mathbf{Y}_t$

$$\mathbf{Y}_t = \boldsymbol{\mu} + \mathbf{X}_t \mathbf{W}_{d\times k}^T + \boldsymbol{\epsilon}_t,$$

for a vector of constants $\boldsymbol{\mu} \in \mathbb{R}^d$, a matrix $\mathbf{W} \in \mathbb{R}^{d\times k}$ and d-dimensional error term random vector $\boldsymbol{\epsilon}_t$. Our first contribution is to separate the tail effect of the error terms and factors that reflects the representation of the original

data in the new basis. It allows for independent assumptions about the profiles of heavy tails of the error term and the original representation.

Secondly, we show how to employ Grouped t-copula into the PPCA framework. The Grouped t-copula allows for a grouped or individual degrees of freedom parameter per marginal of a random vector. It has been explored in [3]. The new component allows the marginal elements of unobserved vectors to have individual or grouped profiles of heavy-tails dependency structures and, consequently, provides greater flexibility in capturing second order characteristics of the data set.

Our next contribution is to combine the described concepts with the flexibility of modelling an asymmetric correlation and heavy- tail dependence in a multivariate setting. We focus on the skewed Student-t distribution from the Generalized Hyperbolic family of distributions as defined and discussed in [4] and comprehensively compared with other families of skewed distributions in [5].

The new PPCA model is based on the particular assumption on the random vectors $\mathbf{X}_t$ and $\boldsymbol{\epsilon}_t$ that are defined as following. Let us denote two mutually independent and identically distributed over time uniform random variables $S_{\epsilon,t}, S_{x,t} \sim \mathcal{U}(0,1)$. For convenience of the notation, we denote d- and k-dimensional random vectors $\mathbf{U}_t$ and $\mathbf{V}_t$, respectively,

$$\mathbf{U}_t = \left(\frac{\chi^{-1}_{\nu^1_\epsilon}(S_{\epsilon,t})}{\nu^1_\epsilon}, \ldots, \frac{\chi^{-1}_{\nu^d_\epsilon}(S_{\epsilon,t})}{\nu^d_\epsilon} \right)_{1\times d} \text{ and } \mathbf{V}_t = \left(\frac{\chi^{-1}_{\nu^1_x}(S_{x,t})}{\nu^1_x}, \ldots, \frac{\chi^{-1}_{\nu^k_x}(S_{x,t})}{\nu^k_x} \right)_{1\times k},$$

for vectors of non-negative real numbers $\boldsymbol{\nu}_\epsilon = \{\nu^1_\epsilon, \ldots, \nu^d_\epsilon\}$ and $\boldsymbol{\nu}_x = \{\nu^1_x, \ldots, \nu^k_x\}$ and χ^{-1}_ν denoting the quantile function of the Chi-square distribution with ν degrees of freedom. Note, that each of the vectors $\mathbf{U}_t$ and $\mathbf{V}_t$ follows a multivariate Gamma distribution, are mutually independent and independent in time. Let us denote d-dimensional and k-dimensional real valued model parameter vectors, $\boldsymbol{\delta}_\epsilon$ and $\boldsymbol{\delta}_x$. The stochastic representation of the d-dimensional error term $\boldsymbol{\epsilon}_t$ and the k-dimensional latent variable $\mathbf{X}_t$ is given by

$$\mathbf{X}_t = \mathbf{V}^{-1}_t \circ \boldsymbol{\delta}_x + \sqrt{\mathbf{V}^{-1}_t} \circ \mathbf{Z}_{x,t} \text{ and } \boldsymbol{\epsilon}_t = \mathbf{U}^{-1}_t \circ \boldsymbol{\delta}_\epsilon + \sqrt{\sigma^2 \mathbf{U}^{-1}_t} \circ \mathbf{Z}_{\epsilon,t},$$

for $\mathbf{Z}_{x,t}$ and $\mathbf{Z}_{\epsilon,t}$ being mutually independent standard normal multivariate variables, k- and d-dimensional respectively, and being independent of $\mathbf{U}_t$ and $\mathbf{V}_t$ (or $S_{x,t}$ and $S_{\epsilon,t}$ likewise). The operator $\circ$ denotes the Hadamard product, that is, for two d-dimensional vectors $\mathbf{a}$ and $\mathbf{b}$, the product of thee vectors results in the d-dimensional vector $\mathbf{a} \circ \mathbf{b} = (a^1 b^1, \ldots, a^d b^d)$.

We comments that this new model for PPCA can easily be reduced to the simpler representation such as Gaussian PPCA and Student-t PPCA if the data reveals such characteristics.

We develop an efficient Expectation-Maximisation (EM) algorithm of [6] that estimates the parameters of this new class of PPCA methods. The framework handles the presence of missing data, and we comment how the procedure can be adjusted to various assumptions about the patterns of missing data.

In addition, we study the robustness of the developed class of the PPCA as defined by [7]. The introduced structural components of the representation of the data generating process increase the flexibility of PPCA frameworks to take into account different features of the data. The introduced models provide estimators of parameters, such as mean and covariance, that contain additional flexibility to accommodate characteristics of the data such as skewness or various patterns of the marginal tail dependence. The proposed estimators are obtained as solutions to the estimation equations that may be seen as 'distorted' in comparison to the equations for the same parameters in the standard PPCA. We argue that this new class of estimators is more robust than the standard PPCA frameworks. We propose to study the characteristics of robustness by the notion of the influence functions and study them numerically in three different ways: by the asymptotic bias of an estimator, the asymptotic variance of an estimator, ie, the precision of the estimation, and by the sensitivity of an estimator to the effect of outliers.

We apply our framework to cryptocurrencies data, and show how the new methodology can be accommodated to guide portfolio construction by measuring market concentration, the potential for diversification or hedging. PCA or PPCA methods can be used to measure market concentration and the potential for diversification. They are often employed to identify highly concentrated assets or to reduce the complexity of large sets of financial instruments by transforming them into a new set of uncorrelated components. Overall results of the covariance decompositions on the set of 20 crypto currencies with the highest market capitalisation in 2018 and 2019 indicated that the majority of the altcoin crypto assets are driven by a common factor that is highly correlated to Bitcoin. This co-moving group of assets is characterized by the highest contribution to the overall variance. The identification of this collinearity can aid the portfolio selection and management as holding only one of these assets provide most of the benefits for the diversification and allows to invested funds in other assets. On the other had, the remaining principal direction indicate uncorrelated assets that can be used for the risk hedging purposes.

References

[1] H Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417–441, 1933.

[2] M. E. Tipping and C. M. Bishop. Probabilistic Principal Component Analysis. *Journal of the Royal Statistical Society, Series B*, 61(3):622–661, 1999.

[3] Stéphane Daul, Filip Lindskog, and Alexander McNeil. The Grouped t-copula with an Application to Credit Risk. *RISK*, 16:73–76, 2003.

[4] O. Barndorff-Nielsen and P. Blaesild. Hyperbolic distributions and ramifications: Contributions to theory and application. In *Statistical Distributions in Scientific Work*, pages 19–44, 1981.

[5] K. Aas and I. H. Haff. The Generalised Hyperbolic Skew Student's t-distribution. *Journal of Financial Econometrics*, 4(2):275–309, 2006.

[6] A. P. Dempster, M. N. Laird, and D. B. Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of Royal Statistical Society, Series B*, 39(1):1–38, 1977.

[7] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and Stahel W. A. *Robust Statistics: The Approach Based on Influence Functions*. Wiley, 1986.