Stochastic Convex Sparse Principal Component Analysis*

Inci M. Baytas[1], Kaixiang Lin[1], Fei Wang[2], Anil K. Jain[1] and Jiayu Zhou[1]
[1]Michigan State University, East Lansing, MI
[2] Cornell University, New York, NY
{baytasin,linkaixi,jain,jiayuz}@msu.edu,few2001@med.cornell.edu

Abstract

Principal Component Analysis (PCA) is a dimensionality reduction and data analysis tool commonly used in many areas. Principal components extracted by the conventional PCA are linear combinations of all the original features. Therefore, we lose the ability to interpret the physical meanings of pricipal components. Sparse PCA has been proposed to improve the interpretability by incorporating sparsity-inducing norms. On the other hand, the sparse PCA has imposed significant computational challenges that prevent it from being applied to explore large-scale datasets. In this paper, we propose a convex sparse principle component analysis (Cvx-SPCA), which leverages a proximal variance reduced stochastic scheme to achieve a geometric convergence rate. We further show that the convergence analysis can be significantly simplified by using a weak condition which allows a broader class of objectives to be applied. The efficiency and effectiveness of the proposed method are demonstrated on a large-scale electronic medical record cohort.

1 Stochastic Convex Sparse Principal Component Analysis

The goal of PCA is to learn a linear transformation such that the learned principle components are the dimensions retaining the most of the variance in the data. Let $\mathbf{X} = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T$ be the normalized covariance matrix for n training data points where each data point is in a d-dimensional feature space. The PCA of computing the top p components can be written as the following optimization problem:

$$\max_{\mathbf{Z}\in\mathbb{R}^{d\times p}} \|\mathbf{XZ}\|_F^2, \quad \text{s.t. } \mathbf{Z}^T\mathbf{Z} = \mathbf{I} \tag{1}$$

where $\mathbf{Z}$ is an orthogonal projection matrix. In this paper, we focus on finding the leading principal component. In the solution of Eq. 1, the principal components (PCs) are linear combinations of all input variables. This means that the columns of $\mathbf{Z}$ matrix, which are called loadings of PCs, are dense. PCA works well, if the interpretation of principal components is not crucial for the application. However, the intepretability is a significant factor, when it comes to many applications such as biomedical informatics. As more and more electronic medical records (EMR) are available of patients, medical researchers are interested in applying various techniques to analyze the EMR data. Typically each feature of EMR data is given by a record/event related to certain diagnosis. When the traditional PCA is applied to the data, these medical features are projected to a low dimensional space, in which each new feature will be the linear combination of all the original features. In this case, it is hard to comprehend the meaning of the new features.

Sparse PCA has been proposed to address this drawback. In sparse PCA, we learn sparse loading vectors which combines only a few of the input variables allowing interpretation of the PCs. In this study, sparse PCA is posed as an ℓ_1 regularized optimization problem. Standard approaches to solve such sparse learning problems are proximal gradient methods, which require the computation of the full gradient at each iteration. Thus they are hardly scalable to large-scale problems with large sample sizes. Therefore, stochastic gradient based methods are preferred in such problems. One major disadvantage of the stochastic gradient descent is the low convergence due to high variance by random sampling.

To tackle the aforementioned challenges in this paper, we propose a novel stochastic convex sparse PCA (Cvx-SPCA) method which is extremely efficient and can handle large-scale datasets. Specifically, we propose to adopt a convex formulation of PCA which provides a strongly convex function. The problem structure in this design allows us to leverage efficient scheme of proximal stochastic gradient with variance reduction (Prox-SVRG) which leads to an exponential (geometric) convergence rate. We also investigate the convergence analysis of Prox-SVRG and present a new proof of the convergence rate which significantly reduces the conditions and assumptions required. As such, we show that the optimization scheme can be applied to a much larger class of problems to obtain the geometric convergence rate. We conducted extensive experiments on both synthetic and real datasets to illustrate the efficiency of the proposed algorithm. The proposed Cvx-SPCA enables us to analyze a large-scale EMR cohort, which is hardly possible by traditional approaches.

*This material is based in part upon work supported by the National Science Foundation under Grant Numbers IIS-1565596 and Office of Naval Research N00014-14-1-0631.