About Recent Work on Sparse PCA with Applications

A short introduction

Yang Xu[1]
yax087@ucsd.edu

[1]Department of Mathematics
University of California, San Diego

March 14, 2019

The Standard PCA

Principal components analysis (PCA) (Jolliffe 1986) is a classical method for the dimension reduction of data in the form of n observations of a vector with p variables.
Consider feature matrix $X \in \mathbb{R}^{n \times p}$, let $X \leftarrow X - \mu$, do SVD on X that

$$X = UDV^T \tag{1}$$

Therefore, $Z = UD$ are principle components (PCs) and the columns of V are corresponding loadings of the principal components.

Problems in standard PCA:

- When p is comparable or dominates n (say, $p > n$), standard PCA is not consistent
- Each principal component is a linear combination of all the original variables and the loadings are typically nonzero, thus it is often difficult to interpret the results (derived PCs)[2]

Sparse Approximations

Zou, Hastie and Tibshirani (2006)[4],base on work of d'Aspremont (2005)[1] showed how PCA can be converted to a (naïve) elastic-net regression problem. They gave the following theorem:

Theorem 1

For each i, denote by $Z_i = U_i D_{ii}$ the ith principal component. Then $\hat{\beta}_{en}$ given by

$$\hat{\beta}_{\text{en}} = \arg\min_{\beta} \|Z_i - X\beta\|^2 + \lambda_2\|\beta\|^2 + \lambda_1\|\beta\|_1 \tag{2}$$

$\hat{V}_i = \hat{\beta}_{\text{en}}/\|\hat{\beta}_{\text{en}}\|, X\hat{V}_i$ are approximations to V_i and principal component.

Remark

Lasso: Need $n \geq k$ where k is a given amount of PCs, produces sparsity.
Elastic-Net: When $p > n$, choose some $\lambda_2 > 0$, then we can include all variables in the fitted model[3].

Sparse Approximations

Theorem 2

For any $\lambda > 0$, suppose we are considering the first k PCs, Let x_i be the ith row vector, $A_{p\times k} = [\alpha_1, \dots, \alpha_k]$ orthonormal, and $B_{p\times k} = [\beta_1, \dots, \beta_k]$, then

$$(\hat{A}, \hat{B}) = \arg\min_{A,B} \sum_{i=1}^{n} \|x_i - AB^T x_i\|^2 + \lambda \sum_{j=1}^{k} \|\beta_j\|^2 + \sum_{j=1}^{k} \lambda_{1,j} \|\beta_j\|_1, s.t. A^T A = I_{k\times k}.$$

Then $\hat{\beta}_j = cV_j, j = 1, 2, \dots, k$ where c is a constant.

It effectively transform the PCA into a regression-type problem. However, this theorem uses lasso to fit A, B, which constrains k.

An Algorithm to Minimize SPCA Criterion

1. B given A: For each j, let $Y_j^* = X\alpha_j$, then each elastic-net estimator

$$\begin{aligned}\hat{\beta}_j &= \arg\min_{\beta} \sum_{i=1}^{n} \|Y_j^* - X\beta_j\|^2 + \lambda\|\beta_j\|^2 + \lambda_{1,j}\|\beta_j\|_1 \\ &= \arg\min_{\beta} (\alpha_j - \beta)^T X^T X(\alpha_j - \beta) + \lambda\|\beta\|^2 + \lambda_{1,j}\|\beta\|_1\end{aligned} \tag{3}$$

2. A given B: Do SVD

$$X^T XB = UDV^T \tag{4}$$

set estimator $\hat{A} = UV^T$.

Algorithm 1 General SPCA Algorithm

1: Let A start at V[,1:k] which are loadings of the first k ordinary PCs.

2: Given a fixed $A = [\alpha_1, \dots, \alpha_k]$, solve the following elastic-net for $j = 1, 2, \dots, k$

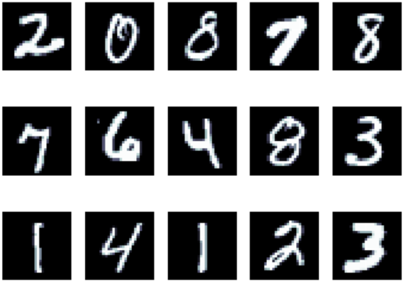
$$\beta_j = \arg\min_{\beta} (\alpha_j - \beta)^T X^T X(\alpha_j - \beta) + \lambda\|\beta\|^2 + \lambda_{1,j}\|\beta\|_1$$

3: For a fixed $B = [\beta_1, \dots, \beta_k]$, compute the SVD of $X^T XB$, then update $A = UV^T$.

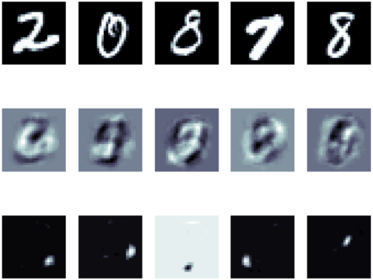
4: Repeat Steps 2-3, until convergence or the maximum number of iterations.

5: Normalization: $\hat{V}_j = \beta_j/\|\beta_j\|, j = 1, \dots, k$.

Handwritten digit recognition. Use the MNIST dataset, $28 \times 28 = 784$ pixels (features) for each picture. Set $n = 100, k = 50$. Perform both PCA and SPCA on the selected data.



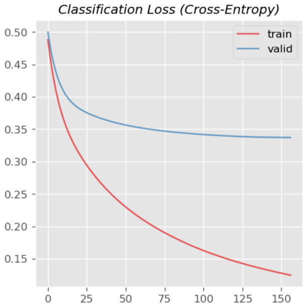
(a) Handwritten digit



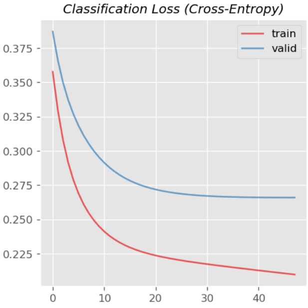
(b) PCA loadings, SPCA loadings

Figure: Comparison between loadings in PCA and SPCA

Numerical Experiment



(a) PCA



(b) SPCA

Figure: Comparison between classification error in PCA and SPCA

Preprocessing Method	Validation Error
PCA	0.337
SPCA	0.266

Reference

Alexandre d'Aspremont, Laurent E Ghaoui, Michael I Jordan, and Gert R Lanckriet.
A direct formulation for sparse pca using semidefinite programming.
In *Advances in neural information processing systems*, pages 41–48, 2005.

Iain M Johnstone and Arthur Yu Lu.
Sparse principal components analysis.
Unpublished manuscript, 7, 2004.

Hui Zou and Trevor Hastie.
Regularization and variable selection via the elastic net.
Journal of the royal statistical society: series B (statistical methodology), 67(2):301–320, 2005.

Hui Zou, Trevor Hastie, and Robert Tibshirani.
Sparse principal component analysis.
Journal of computational and graphical statistics, 15(2):265–286, 2006.