Sparse Principal Component Analysis for Frequency Data

Tobias Bork

Institute for Numerical Simulation

December 9, 2019





Sparse Principal Component Analysis

Tobias Bork

ntroduction

PCA

Fundamentals

parse PCA

Application

PCA

Fundamentals

Sparse PCA

Application

References

Problems in high dimensions:

Time and storage space

Multi-collinearity

Visualizing data set

Curse of dimensionality

▶ **Idea:** Reduce the number of variables while preserving structure in the data

Approach: Feature selection methods

► **Approach:** Feature extraction methods

Central Idea

Sparse Principal Component Analysis

Tobias Bork

Idea

Reduce dimensionality while retaining as much information as possible

- Sequentially identify principal axis of greatest variability
- Represent data set regarding identified principal axis
- Linearly project the data to a space of fewer dimensions
- Yields natural order on principal components

(a) Finding principal axis on a data set

(b) Linear projection of data to a space of fewer dimensions

Sparse Principal Component Analysis

Tobias Bork

ntroduction

CA

Idea

Mathematical Formulation

heorems imits of Usabilit

oparse PCA

ррпсацоп

Mathematical Formulations

Limits of Usability Application

Fundamentals

, pai 50 1 0/1

References

References

Let $\mathbf{X} \in \mathbb{R}^{n \times p}$ be a centered data matrix with n samples and p variables. We find the first principal axis by

$$v_1 = \underset{\|v\|_2 = 1}{\operatorname{arg \, max}} \operatorname{Var}[\mathbf{X}v] = \underset{\|v\|_2 = 1}{\operatorname{arg \, max}} v^T \mathbf{\Sigma} v$$

where $\Sigma = X^T X$ is the sample covariance matrix. We compute the following principal axis successively

$$v_{k+1} = arg \max_{\|v\|=1} v^T \mathbf{\Sigma} v$$

subject to
$$v_{k+1}^T v_l = 0 \quad \forall 1 \le l \le k$$

The new principal components are defined by $Z_i = \mathbf{X}v_i$

The principal axis can also be computed via the eigendecomposition of Σ .

$$\pmb{\Sigma} = \pmb{\mathsf{VLV}}^{\mathsf{T}}$$

where ${\bf L}$ is a diagonal matrix with eigenvalues λ_i and ${\bf V}$ is the matrix of eigenvectors. Closely related is the Singular Value Decomposition (SVD)

$$X = UDV^T$$

where **D** is a diagonal matrix with singular values d_1, \ldots, d_p , **U** a $n \times p$ and **V** a $p \times p$ orthogonal matrix.

Sparse Principal Component Analysis

Tobias Bork

ntroduction

PCA

Idea

Mathematical Formulations

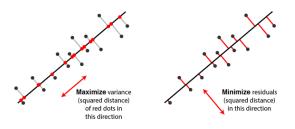
Theorems

Application

unuamentais

......

PCA as a regression problem



Suppose we want to extract the first k principal axis.

$$\hat{\mathbf{V}}_k = \arg\min_{\mathbf{V}_k} \sum_{i=1}^n \left\| x_i - \mathbf{V}_k \mathbf{V}_k^T x_i \right\|^2 + \lambda \sum_{j=1}^k \|\beta_j\|^2$$

subject to $\mathbf{V}_k^T \mathbf{V}_k = \mathbf{I}_{k \times k}$

where x_i is the *i*th row of **X**

Sparse Principal Component Analysis

Tobias Bork

Introduction

PCA

Idos

Mathematical Formulations

Application

undamentais

Аррисаціон



Maska

Mathematical Formulation

Theorems

Application

Fundamentals

Sparse PCA

References

Succes of PCA is due to the following two important optimal properties

- 1. Principal Components sequentially capture the maximum variability (among the columns of X, thus guaranteeing minimal information loss)
- Principal Components are uncorrelated, (so we can talk about one principal component without referring to others)

Theorems

Theorem (Eckart-Young-Mirsky-Theorem)

Let $\widehat{\mathbf{A}}^* = \mathbf{U}_1 \mathbf{D}_1 \mathbf{V}_1^{\top}$ be the truncated singular value decomposition. Then $\widehat{\mathbf{A}}^*$ solves the matrix rank approximation problem

$$\min_{\mathrm{rank}(\widehat{\mathbf{A}}) \leq r} \|\mathbf{A} - \widehat{\mathbf{A}}\|_F = \|\mathbf{A} - \widehat{\mathbf{A}}^*\|_F = \sqrt{\sigma_{r+1}^2 + \dots + \sigma_m^2}$$

where σ_i are the singular values of **A**.

Theorem

PCA is inconsistent for $p \gg n$.

Sparse Principal Component Analysis

Tobias Bork

ntroduction

PCA

Idea

Mathematical Formulatio

Theorems

Application

Fundamentals

parse PCA

Application

Limits of Usability

- ► Linear Relationship between variables
- Correlation of variables
- Completeness of data set
- Outliers
- ▶ Inconsistency theorem in $p \gg n$ case
- ► Interpreation of principal axis

Sparse Principal Component Analysis

Tobias Bork

Introduction

PC

Idea

Mathematical Formulations

I heorems

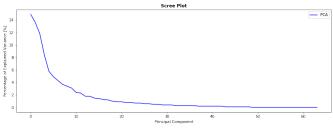
Limits of Usability

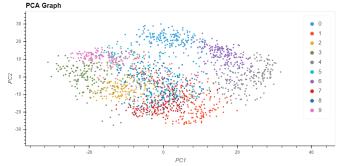
.

undamentais

pplication

Application to handwritten digits





Sparse Principal Component Analysis

Tobias Bork

ntroduction

PCA

Idos

Mathematical Formulation

Theorems

Limits of Usability

Application

Poforoncos



Sparsity inducing Norn

Sparse PCA

Application

deferences

Consider a linear regression model with n observations and p predictors. Let $Y = (y_1, \ldots, y_n)^T$ be the response vector and $\mathbf{X} = [X_1|\cdots|X_p]$. We assume that all the X_j and Y are centered.

The linear regression model has the form

$$f(\mathbf{X}) = \beta_0 + \sum_{j=1}^p X_j \beta_j$$

where the β_j 's are unknown coefficients.

We define the residual sum of squares

$$RSS(\beta) = \sum_{i=1}^{n} (y_i - f(x_i))^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{i=1}^{p} x_{ij} \beta_i)^2 = \|Y - \mathbf{X}\beta\|_2^2$$

$$\hat{eta} = \mathop{\mathsf{arg\,min}}_{eta} \mathit{RSS}(eta)$$

Ridge Regression

Sparse Principal Component Analysis

Tobias Bork

ntroduction

CA

Fundamentals

Regression

. .

TODIAS DOI

 $\hat{\beta}^{\textit{lasso}} = \arg\min_{\beta} \|Y - \mathbf{X}\beta\|_{2}^{2} + \lambda \|\beta\|_{2}^{2} = \sum_{i=1}^{n} (y_{i} - \beta_{0} - \sum_{i=1}^{p} x_{ij}\beta_{j})^{2}$

subject to $\|\beta\|_2^2 \le t$

or equivalently in Lagrangian Form

$$\hat{\beta}^{lasso} = \arg\min_{\beta} \left\{ \frac{1}{2} \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij} \beta_j)^2 + \lambda \|\beta\|_2^2 \right\}$$

LASSO Regression

Sparse Principal Component Analysis

Tobias Bork

ntroduction

PCA

Fundamental:

Regression

Sparsity inducing Norms

Sparse FCA

Application

Elastic Net

Sparse Principal Component Analysis

Tobias Bork

ntroduction

PCA

Fundamental

Regression

Sparsity inducing Norms

Sparse PC/

Application

Sparse PCA

Problem: Principal Components are hard to interpret **Approach:** Require sparse loadings when performing PCA

$$\max v^T \Sigma v$$

subject to
$$\|v\|_{2} = 1$$
, $\|v\|_{0} \le k$

Relaxation:

- a regression framework
- a convex semidefinite programming framework
- ▶ a generalized power method framework
- ▶ an alternating maximization framework
- forward-backward greedy search and exact methods using branch-and-bound techniques
- ► Bayesian formulation framework

Sparse Principal Component Analysis

Tobias Bork

Introduction

PCA

Fundamentals

Sparse PCA

Mathematical Formulation Numerical Solution

F CC .. ----

Mathematical Formulation

We will use a regression framework to derive sparse PCA. We add

$$(\hat{\mathbf{A}}, \hat{\mathbf{B}}) = \underset{\mathbf{A}, \mathbf{B}}{\operatorname{arg\,min}} \sum_{i=1}^{n} \left\| x_i - \mathbf{A} \mathbf{B}^T x_i \right\|^2 + \lambda \sum_{j=1}^{k} \left\| \beta_j \right\|^2 + \sum_{j=1}^{k} \lambda_{1,j} \left\| \beta_j \right\|_{1}^{\text{Adjusted Variant}} \underbrace{\left\| \beta_j \right\|^2 + \sum_{j=1}^{k} \lambda_{1,j} \left\| \beta_j \right\|_{1}^{\text{Adjusted Variant}}}_{\text{References}}$$

subject to
$$\mathbf{A}^T \mathbf{A} = I_{k \times k}$$

- 1: procedure SPCA(A, B)
- 2: $\mathbf{A} \leftarrow \mathbf{V}[1:k]$, the loadings of the first k ordinary principal components
- 3: while not converged do ▷ Definiere Abbruchkriterium
- 4: Given a fixed $\mathbf{A} = [\alpha_1, \dots, \alpha_k]$, solve the elastic net problem

$$\beta_{j} = \operatorname*{arg\,min}_{\beta} \left\| \mathbf{X} \alpha_{j} - \mathbf{X} \beta \right\|^{2} + \lambda \left\| \beta \right\|^{2} + \lambda_{1,j} \left\| \beta \right\|_{1}$$

5: For a fixed $\mathbf{B} = [\beta_1, \dots, \beta_k]$, compute the SVD of

$$\mathbf{X}^T \mathbf{X} \mathbf{B} = \mathbf{U} \mathbf{D} \mathbf{V}^T$$

- 6: $\mathbf{A} \leftarrow \mathbf{U}\mathbf{V}^T$
- 7: end while
- 8: $\hat{V}_j = \frac{\beta_j}{\|\beta_i\|}$ for $j = 1, \dots, k$
- 9: end procedure

Sparse Principal Component Analysis

Tobias Bork

troduction

PCA

Fundamental

parse PCA

Mathematical Formulat

Numerical Solution
Adjusted Variances

p ¿¿ n case

- Sparse Principal Component Analysis
 - Tobias Bork

- References

- Beamer Paket http://latex-beamer.sourceforge.net/
- User's Guide to the Beamer
- DANTE e.V. http://www.dante.de