ACP sparse et group-sparse.

Marie Chavent [1,2] Guy Chavent [3]

[1]Centre Inria de l'université de Bordeaux

[2]Université de Bordeaux

[3]Centre Inria de Paris

6ème journée de Statistique Mathématique
12/01/23

Numerical data matrix A of rank r :

- n observations
- p variables centered or standardized

The aim for Principal Component Analysis :

- find $m \leq r$ vectors $\mathbf{v}_j$ in $\mathbb{R}^p$, $j = 1, \ldots, m$ (loadings),
- to build non correlated variables $\mathbf{y}_j = A\mathbf{v}_j$ (principal components) explaining the maximum variance of A.

The aim of sparse and group-sparse PCA :

- build sparse and group-sparse loading vectors $\mathbf{v}_j$.
- select variables or groups of variables important to build $\mathbf{y}_j$.

Example : data simulated from the model of [5].

- $m = 2$ components, $p = 10$ variables.
- The are variables structured into three groups of size 4, 4, 2.

True loadings

v1	v2
0.422	0.000
0.422	0.000
0.422	0.000
0.422	0.000
0.000	0.489
0.000	0.489
0.000	0.489
0.000	0.489
0.380	-0.147
0.380	0.147

PCA loadings

v1	v2
0.395	0.244
0.338	0.208
0.446	0.027
0.359	0.031
-0.107	0.513
-0.183	0.532
-0.143	0.336
-0.015	0.353
0.417	-0.212
0.402	0.260

Sparse PCA loadings

v1	v2
0.423	0.089
0.354	0.067
0.459	0.000
0.361	0.000
-0.001	0.524
-0.082	0.590
-0.058	0.390
0.000	0.381
0.397	-0.229
0.433	0.125

Group-sparse PCA loadings

v1	v2
0.446	0.000
0.386	0.000
0.464	0.000
0.389	0.000
0.000	0.493
0.000	0.596
0.000	0.453
0.000	0.429
0.360	-0.098
0.395	0.039

$\Rightarrow$ Sparse loadings are not necessarily orthogonal.

Loadings formulation of PCA :

$$\begin{aligned} \max_{\mathbf{v}_j \in \boldsymbol{R}^p} \quad & \|A\mathbf{v}_j\|^2 \\ \text{s.t.} \quad & \|\mathbf{v}_j\| = 1, \\ & \mathbf{v}_i^T \mathbf{v}_j = 0, \quad \forall 1 \leq i < j. \end{aligned} \tag{1}$$

Components formulation of PCA :

$$\begin{aligned} \max_{\mathbf{u}_j \in \boldsymbol{R}^n} \quad & \|A^T\mathbf{u}_j\|^2 = \max_{\mathbf{u}_j \in R^n} \left(\max_{\|\mathbf{v}_j\| \leq 1} \left\langle A^T \mathbf{u}_j, \mathbf{v}_j \right\rangle \right)^2 = \max_{\|\mathbf{v}_j\| \leq 1} \max_{\mathbf{u}_j \in R^n} \left\langle A\mathbf{v}_j, \mathbf{u}_j \right\rangle^2 \\ \text{s.t.} \quad & \|\mathbf{u}_j\| = 1, \\ & \mathbf{u}_i^T \mathbf{u}_j = 0, \quad \forall 1 \leq i < j. \end{aligned} \tag{2}$$

$\Rightarrow$ Used in sparse PCA to handle non-orthogonal sparse loadings.

The solution is given by the singular value decomposition of A :

$$A = U\Sigma V^T$$

- $U = [\mathbf{u}_1 \dots \mathbf{u}_r]$: left singular vectors (eigenvectors of AA^T),
- $V = [\mathbf{v}_1 \dots \mathbf{v}_r]$: right singular vectors (eigenvectors of $A^T A$),
- $\Sigma = \mathrm{diag}(\sigma_1 \dots \sigma_r)$: singular values (square roots of the eigenvalues of AA^T and $A^T A$).

$\Rightarrow$ $\mathbf{u}_j = \frac{A\mathbf{v}_j}{\|A\mathbf{v}_j\|} = \frac{\mathbf{y}_j}{\sigma_j}$ is the standardized principal component,

$\Rightarrow$ The part of the variance of A explained by $Y = [\mathbf{y}_1 \dots \mathbf{y}_m]$ is :

$$\|Y\|_F^2 = \sum_{j=1}^{m} \|\mathbf{y}_j\|^2 = \sigma_1^2 + \dots \sigma_m^2.$$

Component formulation to find the first loading vector $\mathbf{v_1}$:

$$\max_{\|\mathbf{u}\|=1} \|A^T\mathbf{u}\|^2 = \max_{\|\mathbf{u}\|=1} \left(\max_{\|\mathbf{v}\|\leq 1} \left\langle A^T\mathbf{u}, \mathbf{v} \right\rangle \right)^2 \tag{3}$$

Sparse inner maximisation loop ([2],[3]) :

$$\max_{\|\mathbf{v}\|\leq 1} \left(\left\langle A^T\mathbf{u}, \mathbf{v} \right\rangle - \gamma\|\mathbf{v}\|_1 \right). \tag{4}$$

The solution is :

$$\mathbf{v_1} = \frac{S_\gamma(A^T\mathbf{u})}{\|S_\gamma(A^T\mathbf{u})\|}$$

where S_γ is the coordinate-wise **soft-thresholding operator** :

$$(S_\gamma(\mathbf{w}))_i = \frac{w_i}{|w_i|}(|w_i| - \gamma)_+.$$

The maximum of (4) is $\|S_\gamma(A^T\mathbf{u})\|$.

Maximisation of a convex function over a compact :

$$\max_{\|\mathbf{u}\|\leq 1} \underbrace{\|S_\gamma(A^T\mathbf{u})\|^2}_{F(\mathbf{u})}$$

A solution $\mathbf{u}$ can be computed iteratively :

$$\begin{aligned}\mathbf{u}^{(k+1)} &= \arg\max_{\|\mathbf{x}\|\leq 1} \left(F(\mathbf{u}^{(k)}) + \left\langle \nabla F(\mathbf{u}^{(k)}),\ \mathbf{x} - \mathbf{u}^{(k)} \right\rangle\right) \\ &= \arg\max_{\|\mathbf{x}\|\leq 1} \left\langle \nabla F(\mathbf{u}^{(k)}),\ \mathbf{x} \right\rangle .\end{aligned}$$

$$\Rightarrow \mathbf{u}^{(k+1)} = \frac{\nabla F(\mathbf{u}^{(k)})}{\|\nabla F(\mathbf{u}^{(k)})\|}.$$

$\Rightarrow \nabla F(\mathbf{u}) = 2AS_\gamma(A^T\mathbf{u})$ gives :

$$\mathbf{u}^{(k+1)} = \frac{A\ S_\gamma\ \left(A^T\mathbf{u}^{(k)}\right)}{\|A\ S_\gamma\ \left(A^T\mathbf{u}^{(k)}\right)\|}.$$

Algorithm 1 Algorithm to compute the first sparse loading vector $\mathbf{v}_1$.

Choose randomy $\mathbf{u} \in \boldsymbol{R}^n$ such that $\|\mathbf{u}\| = 1$

repeat

$\mathbf{w} \leftarrow S_\gamma(A^T\mathbf{u})$

$\mathbf{u} \leftarrow \frac{A\ \mathbf{w}}{\|A\ \mathbf{w}\|}$

until convergence of $\mathbf{u}$

$\mathbf{v}_1 \leftarrow \frac{\mathbf{w}}{\|\mathbf{w}\|}$

The next sparse loading vectors are computed with same algorithm by deflation :

$$\text{Set } A_0 = A\ ,\ \mathbf{v}_0 = 0\ ,\ \text{and compute, for } j = 1 \ldots m :$$

$$A_j = A_{j-1}(I_p - \mathbf{v}_{j-1}\mathbf{v}_{j-1}^T)$$

$$\mathbf{v}_j = \arg \max_{\|\mathbf{u}_j\|=1} \left(\max_{\|\mathbf{v}_j\| \leq 1} \left(\langle A\mathbf{v}_j, \mathbf{u} \rangle - \gamma_j \|\mathbf{v}_j\|_1 \right) \right)^2$$

Deflation approaches in sparse PCA.

- Shen & Huang, 2008 [5]
- D'Aspremont, Bach & Ghaoui, 2008 [2]
- Mackey 2009 [4] (R package nsprcomp)

Block approaches compute all the sparse loading vectors simultaneously :

- Zou, Hastie & Tibshirani, 2006 [6] (R package elasticnet)
- Journée & al., 2010 [3]

Block components formulation of PCA :

$$\max_{U^TU=I_m} \sum_{j=1}^{m} \|A^T\mathbf{u}_j\|^2 = \max_{U^TU=I_m} \sum_{j=1}^{m} \left(\max_{\|\mathbf{v}_j\|\leq 1} \langle A^T\mathbf{u}_j, \mathbf{v}_j\rangle \right)^2 \tag{5}$$

$$= \max_{\substack{\|\mathbf{v}_j\|\leq 1 \\ j=1,\ldots,m}} \underbrace{\max_{U^TU=I_m} \sum_{j=1}^{m} \langle A\mathbf{v}_j, \mathbf{u}_j\rangle^2}_{optvar(AV)} .$$

with block unknown :

$$V = [v_1 \ \ldots v_m] \in \boldsymbol{R}^{p\times m}$$

$$U = [u_1 \ldots u_m] \in \boldsymbol{R}^{n\times m} \ .$$

Variance of A explained by non necessarily orthogonal principal components.

- Let $Y = [\mathbf{y}_1, \dots, \mathbf{y}_m]$ be the matrix of the m principal components $\mathbf{y}_j = A\mathbf{v}_j$.
- Let $Y' = [\mathbf{y}'_1, \dots, \mathbf{y}'_m]$ be the matrix of the m adjusted components obtained by projection of the $\mathbf{y}_j$ on the j-th axis of an orthonormal basis U :

$$\mathbf{y}'_j = \left\langle \mathbf{y}_j, \mathbf{u}_j \right\rangle \mathbf{u}_j.$$

Hence the projected explained variance of Y estimated with U is :

$$expVar(Y) = \sum_{j=1}^{m} \underbrace{\left\langle \mathbf{y}_j, \mathbf{u}_j \right\rangle^2}_{\|\mathbf{y}'_j\|^2}.$$

- If $U = Q$ of the QR decomposition Y, $expVar(Y)$ is the adjusted variance [6].
- If $U = \arg \max\limits_{U^T U = I_m} \sum_{j=1}^{m} \langle \mathbf{y}_j, \mathbf{u}_j \rangle^2$, $expVar(Y)$ is the optimal variance [1].

The solution of the block PCA optimisation problem is not unique :

- The maximum is achieved for all orthogonal rotation of the the SVD solution.
- The introduction of weights is known to drive the optimization to the SVD solution.

⇒ The weighted block formulation :

$$\max_{U^T U = I_m} \sum_{j=1}^{m} \mu_j^2 \|A^T \mathbf{u}_j\|^2 = \max_{U^T U = I_m} \|A^T U N\|_F^2$$

where

$$\mu_1 > \mu_2 > \cdots > \mu_m$$

and

$$N = \mathrm{diag}(\mu_1, \dots, \mu_m).$$

Block weighted component formulation with inner maximisation :

$$\max_{U^TU=I_m} \sum_{j=1}^{m} \mu_j^2 \|A^T\mathbf{u}_j\|^2 = \max_{U^TU=I_m} \sum_{j=1}^{m} \mu_j^2 \left(\max_{\|\mathbf{v}_j\|\leq 1} \langle A^T\mathbf{u}_j, \mathbf{v}_j \rangle \right)^2$$

Sparse inner maximisation loops [3] :

$$\sum_{j=1}^{m} \mu_j^2 \left(\max_{\|\mathbf{v}_j\|\leq 1} \left(\langle A^T\mathbf{u}_j, \mathbf{v} \rangle - \gamma_j \|\mathbf{v}_j\|_1 \right) \right)^2 \qquad (6)$$

$\Rightarrow$ m separated problems (4) whose solution are :

$$\mathbf{v}_j = \frac{S_{\gamma_j}(A^T\mathbf{u}_j)}{\|S_{\gamma_j}(A^T\mathbf{u}_j)\|}, \ \forall j = 1, \ldots, m$$

$\Rightarrow$ The maximum of (6) is

$$\sum_{j=1}^{m} \mu_j^2 \|S_{\gamma_j}(A^T\mathbf{u}_j)\|^2 = \|\mathbf{S}_\gamma(A^TU)\, N\|_F^2,$$

with $\mathbf{S}_\gamma$ the column-wise soft-thresholding operator.

Maximisation of a convex function over a compact :

$$\max_{U^T U = I_m} \underbrace{\|\mathbf{S}_\gamma(A^T U) N\|_F^2}_{F(U)},$$

$\Rightarrow$ A solution U can be computed iteratively : :

$$\begin{aligned} U^{(k+1)} &= \arg \max_{X^T X = I_m} \left(F(U^{(k)}) + \left\langle \nabla F(U^{(k)}),\ X - U^{(k)} \right\rangle_F \right) \\ &= \arg \max_{X^T X = I_m} \left\langle \nabla F(U^{(k)}),\ X \right\rangle_F . \end{aligned}$$

$\Rightarrow$ $U^{(k+1)} = \mathrm{polar}\left(\nabla F(U^{(k)}) \right).$

$\Rightarrow$ $\nabla F(U) = 2A\mathbf{S}_\gamma(A^T U) N^2$ gives : :

$$U^{(k+1)} = \mathrm{polar}\left(A\mathbf{S}_\gamma(A^T U^{(k)}) N^2 \right).$$

Algorithm 2 Block algorithm to compute m sparse loading vectors

Choose randomy $U \in \mathbb{R}^{n \times m}$ such that $U^T U = I_m$

repeat

$W \leftarrow \mathbf{S}_\gamma \left(A^T U\right)$

$U \leftarrow \mathrm{polar}\left(AWN^2\right)$

until convergence of U

$V \leftarrow [\mathbf{v}_1, \ldots, \mathbf{v}_m]$ with $\mathbf{v}_j = \frac{\mathbf{w}_j}{\|\mathbf{w}_j\|}$.

Interpretation of the criterion :

$$\max_{U^T U = I_m} \sum_{j=1}^{m} \left(\max_{\|\mathbf{v}_j\| \leq 1} \left(\langle A\mathbf{v}_j, \mathbf{u} \rangle - \gamma_j \|\mathbf{v}_j\|_1 \right) \right)^2 \tag{7}$$

1. Find $\mathbf{v}_j$ which gives a compromise between a component $\mathbf{y}_j = A\mathbf{v}_j$ close to the direction of the component u_j and sparse loadings $\mathbf{v}_j$.
2. Find the orthogonal components $\mathbf{u}_j$ which give the best compromises.

Other interpretation : (7) writes :

$$\max_{\substack{\|\mathbf{v}_j\| \leq 1 \\ j=1,\dots,m}} \underbrace{\max_{U^T U = I_m} \sum_{j=1}^{m} \left(\left(\langle A\mathbf{v}_j, \mathbf{u}_j \rangle - \gamma_j \|\mathbf{v}_j\|_1 \right)_+ \right)^2}_{optvar_\gamma(AV)}$$

Find the m sparse loading vectors $\mathbf{v}_j$ such that the non necessarily orthogonal principal components $\mathbf{y}_j = A\mathbf{v}_j$ explain the maximum of the variance of A.

The variables are now structured in G groups.

- The loading vector $\mathbf{v}$ writes :

$$\mathbf{v}^T = (\mathbf{v}_1^T \dots \mathbf{v}_i^T \dots \mathbf{v}_G^T)$$

where $\mathbf{v}_i \in \boldsymbol{R}^{p_i}$ is the loading vector of the p_i variables in group i.

- The ℓ_1-norm of $\mathbf{v}$ extends group ℓ_1-norm :

$$\|v\|_1 = \sum_{i=1}^{G} \|\mathbf{v}_i\|.$$

to promote the apparition of zeros for some groups of variables.

- The soft-thresholding of $\mathbf{w}^T = (\mathbf{w}_1^T, \dots, \mathbf{w}_i^T, \dots, \mathbf{w}_G^T)$ extends to group-wise soft-thresholding :

$$(S_\gamma(\mathbf{w}))_i = \frac{w_i}{\|w_i\|}(\|w_i\| - \gamma)_+.$$

In order to take the size of the groups into account :

- The group ℓ_1-norm writes :

$$\|v\|_1 = \sum_{i=1}^{G} \sqrt{p_i}\, \|\mathbf{v}_i\|.$$

- The group-wise soft-thresholding writes :

$$(S_\gamma(\mathbf{w}))_i = \frac{w_i}{\|w_i\|}(\|w_i\| - \sqrt{p_i}\,\gamma)_+.$$

Finally, Algorithm 2 applies with this group-wise soft-thresholding operator to compute group-sparse PCA ([1])

Implementation in the R package `sparsePCA`[1] :

```
groupsparsePCA(A, m, lambda, index = 1:ncol(A),
               block = 1, mu = 1/1:m,
               center = TRUE, scale = TRUE)
```

- `A` is the data matrix,
- `m` is the number of components,
- `lambda` is the vector of the *m* reduced sparsity parameters sparsité λ_j,
- `index` defines the groups of variables,
- `block` defines the approach (`block=0` for deflation and `block=1` for block).
- `mu` is the vector of weights used in the block approach.

1. https://github.com/chavent/sparsePCAhttps ://github.com/chavent/sparsePCA

Comparison of the deflation and block approaches.

Data simulated with the eigenvalues $(200, 180, 150, 130, 1 \dots 1)$ and the following loadings :

True loadings

v1	v2	v3	v4
6	0	0	6
-6	0	0	6
6	0	0	6
-6	0	0	6
0	12	12	0
0	12	12	0
0	-12	12	0
0	-12	12	0
-5	8	0	5
-5	8	0	-5
5	8	0	5
5	8	0	-5
4	0	0	-10
4	0	0	-10
4	0	0	-10
4	0	0	-10
8	5	8	5
8	5	-8	5
8	-5	8	5
8	-5	-8	5

In this model :

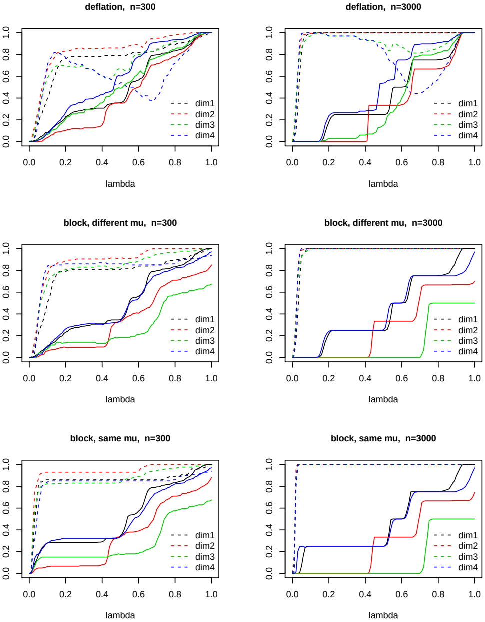
- $p = 20$ variables,
- $m = 4$ components,
- $G = 5$ groups of variables of size 4.

1. Simulation of :
   - 100 matrices A with $n = 300$ observations,
   - 100 matrices A with $n = 3000$ observations.
2. Perform for each matrix A the group-sparse loading matrix V with the deflation approach, the block, same mu approach and the block different mu approach and a grid of reduced sparsity parameters :

   $$\lambda = \lambda_1 = \cdots = \lambda_4 \ ,$$

   with λ varying from 0 to 1 by steps of 0.01.
3. Compare the three approaches with :
   - true positive rate (tpr) : proportion of 0 in V_{true} set to 0 in V,
   - false positive rate (fpr) : proportion of non zéros in V_{true} set to 0 in V.

Mean true positive rates (dotted lines) and false positive rates (full lines).



Sparse PCA of a mixture of numerical and categorical variables. [2]

The Statlog Heart Data Set has $n = 270$ patients described on 7 numerical variables and 6 categorical variables. The variable 'heart disease' is used as illustrative.

Table – Mixed data

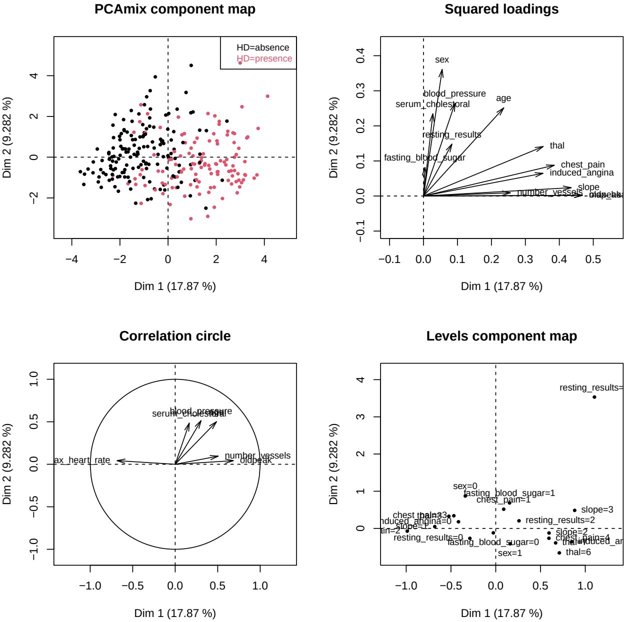
age	oldpeak	sex	slope
70	2.4	1	2
67	1.6	0	2
57	0.3	1	1
64	0.2	1	2
74	0.2	0	3

Table – Recoded numerical data

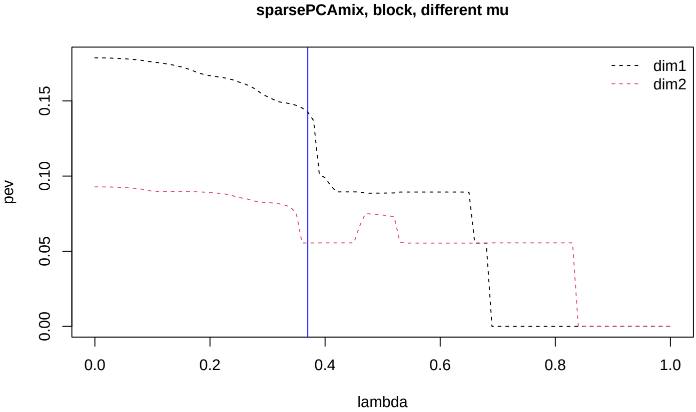
age	oldpeak	sex=0	sex=1	slope=1	slope=2	slope=3
70	2.4	0	1	0	1	0
67	1.6	1	0	0	1	0
57	0.3	0	1	1	0	0
64	0.2	0	1	0	1	0
74	0.2	1	0	0	0	1

The function `sparsePCAmix` uses `groupsparsePCA` with 13 groups (7 groups of size 1 and 6 groups of size = number of levels).

2. https://chavent.github.io/sparsePCA/mixed_data.html


PCAmix component map
HD=absence
HD=presence
Dim 1 (17.87 %)
Dim 2 (9.282 %)
Squared loadings
sex
blood_pressure
serum_cholestoral
age
resting_results
thal
fasting_blood_sugar
chest_pain
induced_angina
slope
number_vessels
Correlation circle
number_vessels
oldpeak
Levels component map
resting_results=
sex=0
fasting_blood_sugar=1
chest_pain=1
slope=3
resting_results=2
slope=2
resting_results=0
fasting_blood_sugar=0
sex=1
thal=6

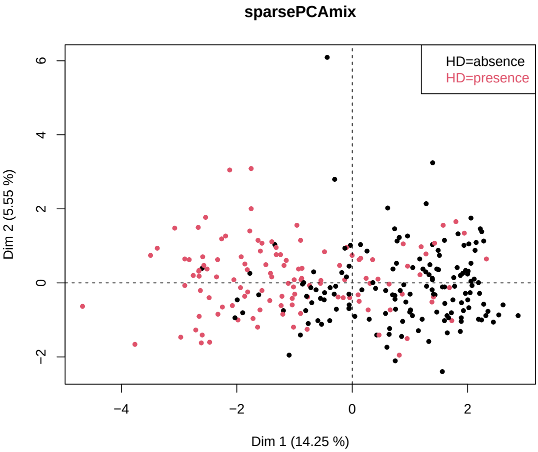
sparsePCAmix with $m = 2$ components.



$\Rightarrow$ Choice $\lambda = \lambda_1 = \lambda_2 = 0.37$.

Table – sparse loadings

	dim 1	dim 2
age	0.000	0
blood_pressure	0.000	0
serum_cholestoral	0.000	1
max_heart_rate	0.431	0
oldpeak	-0.549	0
number_vessels	0.000	0
sex=0	0.000	0
sex=1	0.000	0
chest_pain=1	-0.003	0
chest_pain=2	0.068	0
chest_pain=3	0.053	0
chest_pain=4	-0.117	0
fasting_blood_sugar=0	0.000	0
fasting_blood_sugar=1	0.000	0
resting_results=0	0.000	0
resting_results=1	0.000	0
resting_results=2	0.000	0
induced_angina=0	0.124	0
induced_angina=1	-0.124	0
slope=1	0.283	0
slope=2	-0.222	0
slope=3	-0.061	0
thal=3	0.104	0
thal=6	-0.013	0
thal=7	-0.090	0



sparsePCAmix
HD=absence
HD=presence
Dim 2 (5.55 %)
Dim 1 (14.25 %)
−4
−2
0
2
6
4

Marie Chavent and Guy Chavent.
Optimal Projected Variance Group-Sparse Block PCA.
https://hal.inria.fr/hal-03125264, January 2021.

Alexandre d'Aspremont, Francis Bach, and Laurent El Ghaoui.
Optimal solutions for sparse principal component analysis.
Journal of Machine Learning Research, 9(Jul) :1269–1294, 2008.

Michel Journée, Yurii Nesterov, Peter Richtárik, and Rodolphe Sepulchre.
Generalized power method for sparse principal component analysis.
Journal of Machine Learning Research, 11(Feb) :517–553, 2010.

Lester W Mackey.
Deflation methods for sparse pca.
In *Advances in neural information processing systems*, pages 1017–1024, 2009.

Haipeng Shen and Jianhua Z Huang.
Sparse principal component analysis via regularized low rank matrix approximation.
Journal of multivariate analysis, 99(6) :1015–1034, 2008.

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

Merci de votre attention

Simulation of a matrix A of dimension $n \times p$ from a SVD model :

1. The $m \leq p$ eigenvectors $v_1, \ldots, v_m$ (right singular vectors) and the m eigenvalues $\sigma_1^2 \ldots \sigma_p^2$ of a covariance matrix are given in input and the covariance matrix C is built with :

   $$C = V_{true} \Sigma_{true}^2 V_{true}^T$$

   where :

   - $\Sigma_{true}^2 = \mathrm{diag}(\sigma_1^2 \ldots \sigma_p^2)$ is the diagonal matrix of the eigenvalues,
   - V_{true} is the matrix of the eigenvalues, obtained using the following QR decomposition=

     $$[v_1, \ldots, v_m, U] = V_{true} R,$$

     where U of dimension $p \times (p - m)$ is drawn randomly from a $U(0, 1)$ distribution.

2. n observations are randomly drawn from a $N(0_p, C)$ distribution to get the data matrix A.

The group-soft thersholing operator $S_\gamma(A^T U)$

- The matrix a_i of dimension $n \times p_i$ contains the data of group i and

$$A = \left[a_1 \dots a_i \dots a_p\right] .$$

- The vector $a_i^T u_j \in \boldsymbol{R}^{p_i}$ contains correlations between the p_i variables of group i end the jth normalized component u_j and

$$A^T U = \begin{pmatrix} a_1^T u_1 & \dots & a_1^T u_m \\ \vdots & & \vdots \\ a_G^T u_1 & \dots & a_G^T u_m \end{pmatrix}$$

- The vector $t_{ij} = S_{\gamma_j}(a_i^T u_j)$ is obtained by soft thresholding of the vector $a_i^T u_j$ i.e. the vector is set to 0 if its norm is smaller than γ_j and its length is reduced of γ_j otherwise :

$$t_{ij} = \frac{a_i^T x_j}{\|a_i^T x_j\|}[\|a_i^T x_j\| - \gamma_j]_+ \in \boldsymbol{R}^{p_i} .$$

- Finally

$$S_\gamma(A^T U) = \begin{pmatrix} t_{11} & \dots & t_{1m} \\ \vdots & & \vdots \\ t_{G1} & \dots & t_{G1} \end{pmatrix} = T$$

Each sparsity parameter γ_j needs to be fitted to the norm of the vector $A^T u_j$ it is in charge of thresholding. This norm is simply estimated by its initial value

$$\|A^T u_j^0\| = \sigma_j,$$

where u_j^0 is the jth left singular vectors of A.

Moreover, it can be shown all the loading vector v_j is null if

$$\gamma_j \geq \gamma_{max}$$

with

$$\gamma_{max} \stackrel{\text{def}}{=} \max_{i=1\dots p} \|a_i\|_2$$

and $\|a_i\|_2$ the first singular value of a_i.

A nominal maximum sparsity parameters $\gamma_{j,max}$ is then defined for each dimension :

$$\gamma_{j,max} = \frac{\sigma_j}{\sigma_1} \gamma_{max},$$

and finaly the reduced sparsity parameters λ_j are chosen in $[0, 1]$ with :

$$\lambda_j = \gamma_j / \gamma_{j,max} \quad , \quad j = 1 \dots m \; .$$

When no a priori information on the sparsity of the underlying loadings is known, we can use the same reduced parameters λ for all loadings :

$$\lambda = \lambda_1 = \cdots = \lambda_m \; ,$$

and explore the influence of λ by letting it vary from 0 to 1 by steps of $0,01$ for instance.

Power method to find the first eigenvector of $A^T A$:

Choose randomy $\mathbf{v}^{(0)} \in \boldsymbol{R}^p$

repeat

$$\mathbf{v}^{(k+1)} = \frac{A^T A\ \mathbf{v}^{(k)}}{\|A^T A\ \mathbf{v}^{(k)}\|}$$

until $\|\mathbf{v}^{(k+1)} - \mathbf{v}^{(k)}\| < \epsilon$

Power method to find the first eigenvector of AA^T :

Choose randomy $\mathbf{u}^{(0)} \in \boldsymbol{R}^n$

repeat

$$\mathbf{u}^{(k+1)} = \frac{A\ A^T \mathbf{u}^{(k)}}{\|A\ A^T \mathbf{u}^{(k)}\|}$$

until $\|\mathbf{u}^{(k+1)} - \mathbf{u}^{(k)}\| < \epsilon$

Algorithm 3 Block power method to compute the m first loading vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$.

Choose randomy $U \in \mathbb{R}^{n \times m}$ such that $U^T U = I_m$

repeat

$W \leftarrow A^T U$

$U \leftarrow \mathrm{polar}(AWN^2)$

until convergence of U

$V \leftarrow [\mathbf{v}_1, \ldots, \mathbf{v}_m]$ with $\mathbf{v}_j = \frac{\mathbf{w}_j}{\|\mathbf{w}_j\|}$.