Sparse Principal Component Analysis

Tomáš Masák

13.9.2016

Let $\mathbf{X} \in \mathbb{R}^{n \times p}$ be a data matrix and $\widehat{\mathbf{\Sigma}} = \frac{1}{n}\mathbf{X}^\top\mathbf{X}$ be a sample covariance matrix.

For $k = 1, \ldots, p$, let

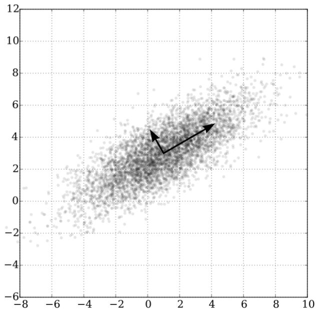
$$\begin{aligned} \widehat{a}_k \quad = \quad & \underset{a}{\arg\max} \quad a^\top \widehat{\mathbf{\Sigma}} a \\ & \text{s.t.} \quad \|a\|_2^2 \leq 1 \ , \\ & \qquad\;\; a^\top \widehat{a}_i = 0 \ , \ i = 1, \ldots, k-1 \ . \end{aligned}$$

Let $\mathbf{X} \in \mathbb{R}^{n \times p}$ be a data matrix and $\widehat{\mathbf{\Sigma}} = \frac{1}{n}\mathbf{X}^\top\mathbf{X}$ be a sample covariance matrix.

For $k = 1, \ldots, p$, let

$$\begin{aligned} \widehat{a}_k \quad = \quad \underset{a}{\text{arg max}} \quad & a^\top \widehat{\mathbf{\Sigma}} a \\ \text{s.t.} \quad & \|a\|_2^2 \leq 1 \ , \\ & a^\top \widehat{a}_i = 0 \ , \ i = 1, \ldots, k-1 \ . \end{aligned}$$

- $\widehat{a}_1, \ldots, \widehat{a}_p$ are *loadings*
- $Y_1 = \mathbf{X}\widehat{a}_1, \ldots, Y_p = \mathbf{X}\widehat{a}_p$ are *principal components* (PCs)
- PCs = uncorrelated standardized linear combinations with the largest possible variance



Zou et al. (2006) showed that loadings can be found as solutions to ridge regression problems

$$\begin{aligned} \widehat{a}_k &= \underset{a}{\arg\max} \quad \|Y_k - \mathbf{X}a\|_2^2 \\ & \quad\text{s.t.} \quad \|a\|_2^2 \leq t \,, \end{aligned}$$

for some t (here completely arbitrary).

Zou et al. (2006) showed that loadings can be found as solutions to ridge regression problems

$$\begin{aligned} \widehat{a}_k \quad = \quad & \underset{a}{\text{arg max}} \quad \|Y_k - \mathbf{X}a\|_2^2 \\ & \text{s.t.} \qquad \|a\|_2^2 \leq t \ , \end{aligned}$$

for some t (here completely arbitrary).

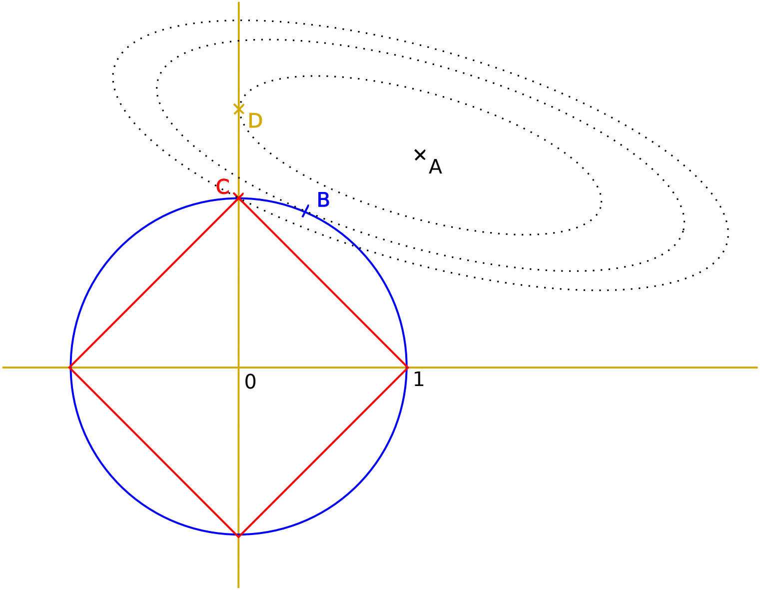
Problems with PCA:

- inconsistency when $p > n$
- poor interpretability – all coefficients non-zero

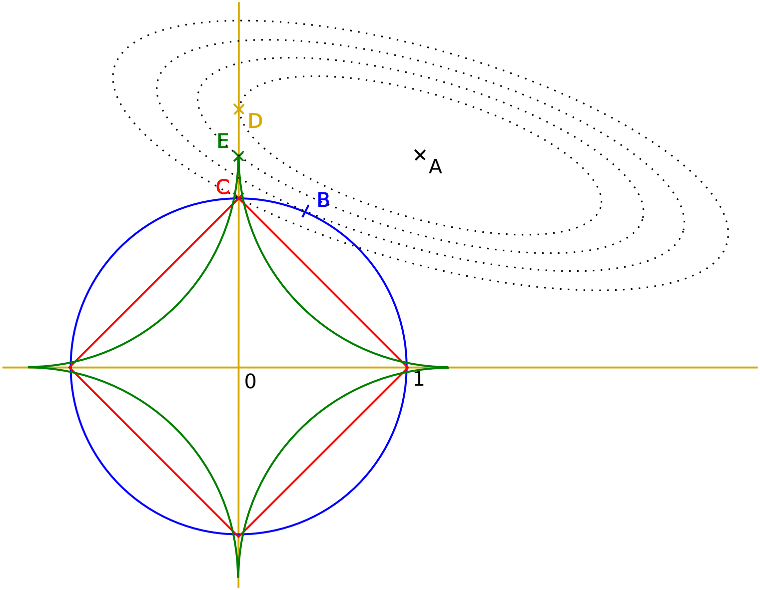
Solution: constraint on cardinality

$$\|a\|_0 \leq s$$

for some s.



$$A: \min_a \|Y_k - \mathbf{X}a\|_2^2$$
$$B: \text{s.t. } \|a\|_2^2 \leq 1$$
$$C: \text{s.t. } \|a\|_1 \leq 1$$
$$D: \text{s.t. } \|a\|_0 \leq 1$$



$$A: \quad \min_a \quad \|Y_k - \mathbf{X}a\|_2^2$$
$$B: \quad \text{s.t.} \quad \|a\|_2^2 \leq 1$$
$$C: \quad \text{s.t.} \quad \|a\|_1 \leq 1$$
$$D: \quad \text{s.t.} \quad \|a\|_0 \leq 1$$

$$E: \quad \min_a \quad \|Y_k - \mathbf{X}a\|_2^2$$
$$\text{s.t.} \quad \sum_{i=1}^{p} w_i |a_i| \leq 1 \ ,$$

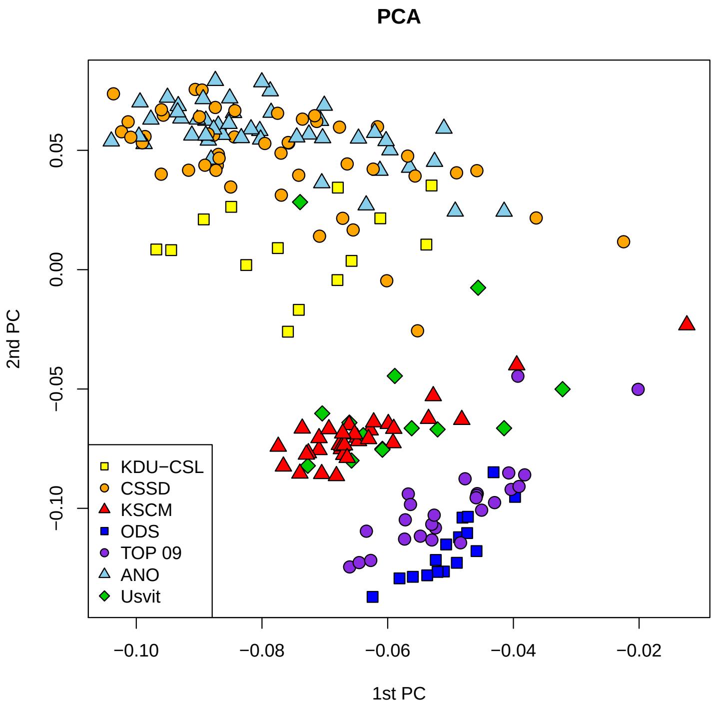
where $w_i = \frac{1}{\tilde{a}_i + \epsilon}$.

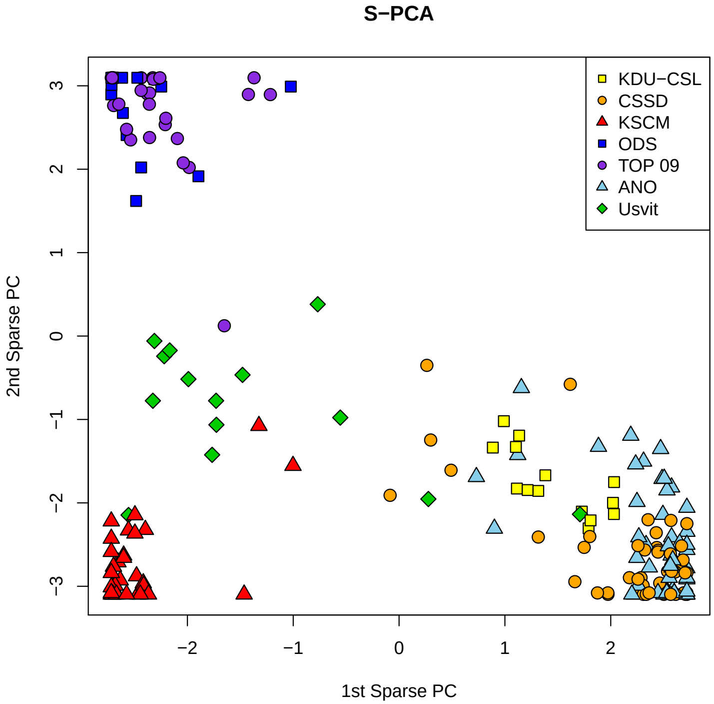
2015 Voting Data

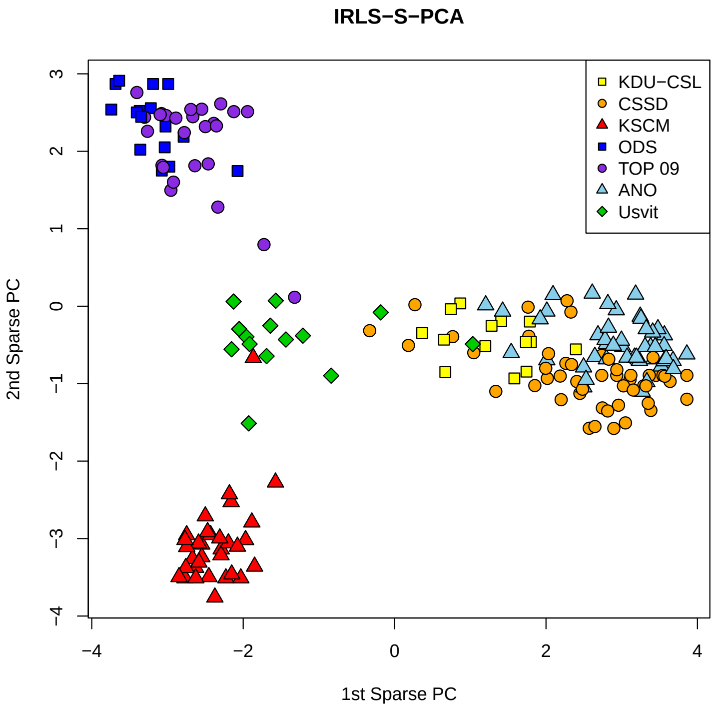
$n = 200$ members of parliament, $p = 1837$ bills voted

- ...
- ...
- Bohuslav Sobotka – vector of votes $x_{Sobotka} \in \{-1, 0, 1\}^p$, and a party affiliation $b \in \{1, \ldots, 5\}$
- ...
- ...

$$x_{Sobotka,j} = \begin{cases} 1, & \text{voted for the } j\text{-th bill} \\ 0, & \text{abstain from voting of the } j\text{-th bill} \\ -1, & \text{voted against the } j\text{-th bill} \end{cases}$$


PCA
2nd PC
1st PC
0.05
0.00
−0.05
−0.10
−0.10
−0.08
−0.06
−0.04
−0.02
KDU−CSL
CSSD
KSCM
ODS
TOP 09
ANO
Usvit


S−PCA
KDU−CSL
CSSD
KSCM
ODS
TOP 09
ANO
Usvit
2nd Sparse PC
1st Sparse PC
3
2
1
0
−1
−2
−3
−2
−1
0
1
2


IRLS−S−PCA
KDU−CSL
CSSD
KSCM
ODS
TOP 09
ANO
Usvit
2nd Sparse PC
1st Sparse PC
3
2
1
0
−1
−2
−3
−4
−4
−2
0
2
4

References

- Amini, A. A., & Wainwright, M. J. (2008). High-dimensional analysis of semidefinite relaxations for sparse principal components. In Information Theory, 2008. ISIT 2008. IEEE International Symposium on (pp. 2454-2458). IEEE.
- d'Aspremont, A., El Ghaoui, L., Jordan, M. I., & Lanckriet, G. R. (2007). A direct formulation for sparse PCA using semidefinite programming. SIAM review, 49(3), 434-448.
- d'Aspremont, A., Bach, F., & Ghaoui, L. E. (2008). Optimal solutions for sparse principal component analysis. The Journal of Machine Learning Research, 9, 1269-1294.
- d'Aspremont, A., Bach, F., & El Ghaoui, L. (2014). Approximation bounds for sparse principal component analysis. Mathematical Programming, 148(1-2), 89-110.
- Berthet, Q., & Rigollet, P. (2013). Optimal detection of sparse principal components in high dimension. The Annals of Statistics, 41(4), 1780-1815.
- Bühlmann, P., & Meier, L. (2008). Discussion of "One-step sparse estimates in nonconcave penalized likelihood models" (auths H. Zou & R. Li). Annals of Statistics 36 1534–1541.
- Bühlmann, P., & Van de Geer, S. (2011). Statistics for high-dimensional data: methods, theory and applications. Springer Science & Business Media.
- Candés, E. J., Wakin, M. B., & Boyd, S. P. (2008). Enhancing sparsity by reweighted ℓ_1 minimization. Journal of Fourier analysis and applications, 14(5-6), 877-905.
- Johnstone, I. M., & Lu, A. Y. (2012). On consistency and sparsity for principal components analysis in high dimensions. Journal of the American Statistical Association.
- Jolliffe, I. (2002). Principal component analysis. John Wiley & Sons, Ltd.
- Leng, C., & Wang, H. (2012). On general adaptive sparse principal component analysis. Journal of Computational and Graphical Statistics.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), 267-288.
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301-320.
- Zou, H., Hastie, T., & Tibshirani, R. (2006). Sparse principal component analysis. Journal of computational and graphical statistics, 15(2), 265-286.
- Zou, H. (2006). The adaptive lasso and its oracle properties. Journal of the American statistical association, 101(476), 1418-1429.
- Zou, H., & Zhang, H. H. (2009). On the adaptive elastic-net with a diverging number of parameters. Annals of statistics, 37(4), 1733.