Penalized Orthogonal Components Regression for Large p Small n Data

Dabao Zhang

Department of Statistics
Purdue University

May 16, 2009

Orthogonal-Components Regression

$$Y = \beta^T X + \epsilon, \qquad X \perp \epsilon,\ E[\epsilon] = 0.$$

Idea

- Sequentially constructs orthogonal components $\varpi_1^T X, \varpi_2 X, \cdots$:

  — $\tilde{X}_1 = X$;

  — $\tilde{X}_k, k \geq 2$ is iteratively built to be orthogonal to $\omega_1^T \tilde{X}_1, \cdots, \omega_{k-1}^T \tilde{X}_{k-1}$;

  — ω_k is obtained to maximize

  $$cov(Y, \omega^T \tilde{X}_k)^T cov(Y, \omega^T \tilde{X}_k), \quad \|\omega\| = 1;$$

  — $\varpi_k^T X = \omega_k^T \tilde{X}_k$;

  — Stop when $Y \perp \tilde{X}_k^T$.

- Orthogonal-Component Regression

$$Y = \sum_k \vartheta_k(\varpi_k^T X) + \epsilon.$$

How to Fit an Orthogonal-Components Regression?

Issue: Need to Estimate $cov(Y, \tilde{X}_k)$!

Partial Least Squares

- Estimate $cov(Y, \tilde{X}_k)$ with the empirical covariance matrix, i.e., $\mathbf{Y}^T\tilde{\mathbf{X}}_k/n$.
- Weakness:

— Lack the ability to select variables;

— Have to choose number of components in the model.

Penalized Orthogonal-Components Regression (POCRE)

Idea

- Find the leading SPARSE eigenvector ω_k of $cov(Y, \tilde{X}_k)^T cov(Y, \tilde{X}_k)$;

  — Obtain $\omega_k = \hat{\gamma}/\|\hat{\gamma}\|$ with

$$(\hat{\alpha}, \hat{\gamma}) = \arg \min_{\alpha,\gamma:\|\alpha\|=1} \left\{ -2\gamma^T \tilde{\mathbf{X}}_k^T \mathbf{Y}\mathbf{Y}^T \tilde{\mathbf{X}}_k \alpha + \|\gamma\|^2 + p_\lambda(\gamma) \right\};$$

- Iteratively solve for $\hat{\alpha}$ and $\hat{\gamma}$,

$$\begin{aligned} \hat{\alpha}(\gamma) &= \arg \min_{\alpha:\|\alpha\|=1} \left\{ -2\gamma^T \tilde{\mathbf{X}}_k^T \mathbf{Y}\mathbf{Y}^T \tilde{\mathbf{X}}_k \alpha \right\} \\ \hat{\gamma}(\alpha) &= \arg \min_{\gamma} \left\{ \|\gamma - \tilde{\mathbf{X}}_k^T \mathbf{Y}\mathbf{Y}^T \tilde{\mathbf{X}}_k \alpha\|^2 + p_\lambda(\gamma) \right\}, \end{aligned}$$

Penalized Orthogonal-Components Regression (POCRE)

Penalization via Empirical Bayes Thresholding

$$\hat{\gamma}(\alpha) = \arg\min_{\gamma} \left\{ \|\gamma - \tilde{\mathbf{X}}_k^T \mathbf{Y}\mathbf{Y}^T \tilde{\mathbf{X}}_k \alpha\|^2 + p_\lambda(\gamma) \right\}$$

- Approximate $\hat{\gamma}(\alpha)$ using the empirical Bayes thresholding by Johnstone and Silverman (2004);
- The empirical Bayes thresholding provides sparsity-adaptive estimate of the eigenvector ω_k.

Simulation Study

Case 1 (High Correlations). $Y = 2\sum_{j=1}^{10} X_j + \sum_{j=101}^{110} X_j + \varepsilon$, where $\varepsilon \sim N(0,1)$, and each block $\{X_{k+1}, \cdots, X_{k+100}\}$ is simulated from an AR(1) process with $\rho = 0.9$, $k = 0, 100, \cdots, 900$.

Case 2 (Mild Correlations). Same as Case 1 except that $\rho = 0.5$.

Case 3 (Clustered Predictors). $Y = 1.5\sum_{j=1}^{30} X_j + \varepsilon$, where $\varepsilon \sim N(0, 15^2)$, and $X_j = Z_1 1_{\{j \le 10\}} + Z_2 1_{\{11 \le j \le 20\}} + Z_3 1_{\{21 \le j \le 30\}} + \xi_j$. Here $Z_1, Z_2, Z_3 \overset{iid}{\sim} N(0,1)$, and $\xi_j \overset{iid}{\sim} N(0, 0.01)$.

Case 4 (Errors in Predictors). $Y = Z_1 + 2Z_2 + Z_3 + \varepsilon$, where $\varepsilon \sim N(0,1)$. Note that $X_j = sign(5.5 - j) Z_1 1_{\{j \le 10\}} + sign(15.5 - j) Z_2 1_{\{11 \le j \le 20\}} + Z_3 1_{\{21 \le j \le 30\}} + \xi_j$, where $Z_1, Z_2, Z_3 \overset{iid}{\sim} N(0,1)$, and $\xi_j \overset{iid}{\sim} N(0,1)$.

Case 5 (Latent-Variable Model). $Y_k = a_k Z_1 + b_k Z_2 + \varepsilon_k$, $1 \le k \le 5$, where $a_1 = a_2 = b_2 = 2$, $b_1 = a_3 = b_3 = -2$, $a_4 = a_5 = 3$, $b_4 = -b_5 = 1.5$, and $\varepsilon_k \overset{iid}{\sim} N(0,1)$. $Z_1 = X_{50} + X_{150} + X_{250} + X_{350} + X_{450} + X_{550}$ and $Z_2 = X_{51} + X_{153} + X_{256} + X_{359} + X_{467} + X_{583}$, where X's are the same as in Case 1 except that $\rho = 0.3$.

Simulation Study

Summary on losses (with standard errors in parentheses)

n	Method	Case 1	Case 2	Case 3	Case 4	Case 5
100	EN	29.80(1.31)	**2.03**(1.53)	103.34(4.35)	**1.45**(0.04)	13.48(1.29)
	Lasso	**0.66**(0.02)	**1.76**(0.10)	**72.12**(4.04)	1.59(0.03)	**12.47**(0.79)
	PCR	310.43(2.62)	123.71(0.44)	228.25(4.48)	2.32(0.03)	281.17(0.47)
	PLS	81.44(1.15)	89.94(0.48)	187.57(3.25)	3.10(0.02)	254.43(0.79)
	POCRE	**1.08**(0.12)	2.97(0.28)	**16.78**(2.90)	**0.86**(0.03)	**2.13**(1.30)
	Ridge	81.60(1.13)	89.71(0.44)	193.90(3.21)	3.09(0.02)	253.18(0.52)
50	EN	39.23(2.09)	52.45(2.65)	**141.90**(7.93)	**2.30**(0.13)	250.51(2.92)
	Lasso	**1.98**(0.13)	**33.24**(1.66)	167.93(9.64)	2.74(0.06)	**234.97**(3.21)
	PCR	308.84(2.87)	124.90(0.57)	378.19(5.39)	3.41(0.05)	282.37(0.61)
	PLS	196.82(2.25)	111.26(0.73)	331.31(4.35)	4.24(0.03)	273.23(0.83)
	POCRE	**2.53**(0.30)	**38.76**(1.79)	**64.94**(8.51)	**1.77**(0.06)	**227.55**(6.55)
	Ridge	192.01(2.26)	110.56(0.53)	333.79(4.45)	4.22(0.03)	269.71(0.62)

Summary on FDR with the best method shown in bold

n	Method	Case 1	Case 2	Case 3	Case 4	Case 5
100	EN	0.9603	0.7260	0.4118	0.7216	0.8452
	Lasso	0.5745	0.7037	0.7931	0.6087	0.8421
	POCRE	**0.1304**	**0.1304**	**0.0000**	**0.0385**	**0.1429**
50	EN	0.9184	0.8365	0.7285	0.8167	0.9622
	Lasso	0.4722	0.6818	0.8222	0.6333	0.8197
	POCRE	**0.1304**	**0.4599**	**0.0657**	**0.5238**	**0.7817**

Acknowledgement

Co-Authors:

Yanzhu Lin
Min Zhang

Thank You!