High-dimensional Mixed Linear Regression

Kiyeon Jeon

University of Texas at Austin

kiyeonj@utexas.edu

June 26, 2018

Overview

Problem Setting : High-dimensional Mixed Linear Regression

- $y_i = \langle x_i, \beta_1^\star \rangle z_i + \langle x_i, \beta_2^\star \rangle (1 - z_i) + w_i$ for $i = 1, \ldots, n$
- $x_i, \beta_i^\star \in \mathbb{R}^d, z_i \in \{0, 1\}, w_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$
- $d \gg n, \beta_k^\star$ is sparse
- **Goal**: infer $\beta_1^\star, \beta_2^\star$ given $\{(x_i, y_i)\}$

Simulation Results

- $y_i = \langle x_i, \beta_1^\star \rangle z_i + \langle x_i, \beta_2^\star \rangle (1 - z_i) + w_i$ for $i = 1, \ldots, n$
- $x_i \in \mathcal{N}(0, I_d), \beta_k^\star \in \mathbb{R}^d, \|\beta_k^\star\|_0 = s, w_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$
- $z_i \in \{0, 1\}, \sum_i z_i = \lfloor np_1 \rfloor$ (p_1 is a fraction of MLR)
- $r = \min_k \{\|supp(\beta_k^\star) \setminus supp(\hat{\beta})\|_0\}$ (r is the failure of target recovery)

$(n, d, s, p1)$	r	$(n, d, s, p1)$	r	$(n, d, s, p1)$	r
(3e2,1e4,10,0.5)	4	(6e2,1e4,20,0.5)	8	(9e2,1e4,30,0.5)	20
(4e2,1e4,10,0.5)	3	(8e2,1e4,20,0.5)	3	(1.2e3,1e4,30,0.5)	0
(5e2,1e4,10,0.5)	0	(1e3,1e4,20,0.5)	0	(1.5e3,1e4,30,0.5)	0

Table: $\sigma^2 = 0.1$

Simulation Results

- $y_i = \langle x_i, \beta_1^\star \rangle z_i + \langle x_i, \beta_2^\star \rangle (1 - z_i) + w_i$ for $i = 1, \dots, n$
- $x_i \in \mathcal{N}(0, I_d), \beta_k^\star \in \mathbb{R}^d, \|\beta_k^\star\|_0 = s, w_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$
- $z_i \in \{0, 1\}, \sum_i z_i = \lfloor np_1 \rfloor$ (p_1 is a fraction of MLR)
- $r = \min_k \{\|supp(\beta_k^\star) \setminus supp(\hat{\beta})\|_0\}$ (r is the failure of target recovery)

(n, d, s, σ^2)	r	(n, d, s, σ^2)	r	(n, d, s, σ^2)	r
(5e2,1e4,10,0.1)	0	(1e3,1e4,20,0.1)	0	(1.5e3,1e4,30,0.1)	0
(5e2,1e4,10,0.12)	1	(1e3,1e4,20,0.12)	1	(1.5e3,1e4,30,0.12)	2
(5e2,1e4,10,0.15)	0	(1e3,1e4,20,0.15)	2	(1.5e3,1e4,30,0.15)	3
(5e2,1e4,10,0.2)	1	(1e3,1e4,20,0.2)	0	(1.5e3,1e4,30,0.2)	6
(5e2,1e4,10,0.3)	5	(1e3,1e4,20,0.3)	4	(1.5e3,1e4,30,0.3)	7

Table: $p_1 = 0.5$

References

Roman Vershynin (2010)
Introduction to the non-asymptotic analysis of random matrices
arXiv preprint arXiv:1011.3027 12(3), 45 – 678.

Nicolas Städler (2010)
1-penalization for mixture regression models
Test, Springer 19, 209–256

Xinyang Yi (2014)
Alternating minimization for mixed linear regression
International Conference on Machine Learning 14, 613–621

The End