High-dimensional Mixed Linear Regression

Kiyeon Jeon

University of Texas at Austin

kiyeonj@utexas.edu

August 13, 2018

Overview

1. Problem Setting
2. Algorithm
3. Simulation
   - Symmetric case
   - Different Proportion
   - Low-dimensional case
   - High-dimensional case

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

Overview

Algorithm 1: Fixed Threshold

Input: $\{(x_i, y_i)\}_{i=1,2,\dots,n}$, $\beta^{(0)}$, $T, G, s, \alpha(< 0.5), \eta$

for $t = 1$ to T **do**
    $J \leftarrow \emptyset$
    $J \leftarrow$ the smallest (αn) index of $|y_i - \langle x_i, \beta^{(t)}\rangle|$
    $\beta^{(t+1)} \leftarrow \text{Update}(X, Y, \beta^{(t)}, J, \eta, s, G)$
end for

Algorithm 2: Reduced Threshold

Input: $\{(x_i, y_i)\}_{i=1,2,\dots,n}$, $\beta^{(0)}$, $T, G, s, \alpha(< 0.5), \eta, K$

for $t = 1$ to T **do**
    $J \leftarrow \emptyset$
    $J \leftarrow$ the smallest $\max\{(n - (t-1)K), n\alpha\}$ index of $|y_i - \langle x_i, \beta^{(t)}\rangle|$
    $\beta^{(t+1)} \leftarrow \text{Update}(X, Y, \beta^{(t)}, J, \eta, s, G)$
end for

Algorithm

Algorithm 3: Reduced Threshold for Mixed Linear Regression

Input:Initial $\beta^{(0)}, T, G, s, C, \{(x_i, y_i)\}_{i=1,2,\ldots,n}$

1: $\beta_1, S_1 \leftarrow$ Threshold $(X, Y, \beta^{(0)}, T, G, s, \alpha, \eta, K)$

2: $\beta_2, S_2 \leftarrow$ Robust Regression $(X, Y, \beta^{(0)}, [n] \setminus S_1)$

Overview

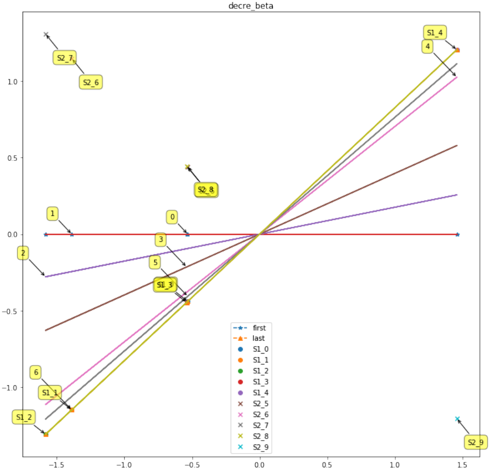
1. Problem Setting
2. Algorithm
3. Simulation
   - Symmetric case
   - Different Proportion
   - Low-dimensional case
   - High-dimensional case

Overview

1. Problem Setting
2. Algorithm
3. Simulation
   - Symmetric case
   - Different Proportion
   - Low-dimensional case
   - High-dimensional case

Symmetric case - x symmetric

$$y_1 = X\beta + e_1, y_2 = -X\beta + e_2$$

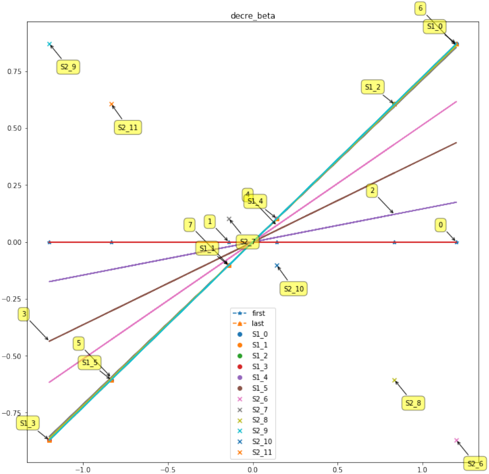


(n, d, σ)	(10,1,0)
Fixed	(5.551e-17 0.05205)
Reduced	(2.776e-17, 4.636e-17)
EM-init	(2.775e-17, 4.635e-17)
EM-rand	(5.551e-17, 0.03268)

Table: (median,mean) over 100 trials(MLR)

Symmetric case - xy symmetric

$y_1 = X\beta + e_1, y_2 = -X\beta + e_2$



(n,d,σ)	(10,1,0)
Fixed	(5.551e-17, 03604)
Reduced	(2.775e-17, 4.561e-17)
EM-init	(2.775e-17, 4.561e-17)
EM-rand	(1.110e-16, 0.03049)

Table: (median,mean) over 100 trials(MLR)

Overview

Simulation Results - Different Proportion

(n, d, p, σ)	(1000,100,0.1,0)	(1000,100,0.2,0)	(1000,100,0.4,0)
Fixed	(3.968e-12, 4.018e-12)	(4.524e-12,4.510e-12)	(5.911e-12,5.8789e-12)
Reduced	(5.590e-12, 5.643e-12)	(6.675e-12,6.680e-12)	(1.253e-11,1.269e-11)
EM-init	(0.03459,0.05530)	(2.638e-17,0.02041)	(4.579e-16,0.01127)
EM-rand	(0.1154,0.1454)	(0.1347,0.1993)	(0.7476,0.6216)

(median,mean) over 100 trials(MLR) - Grad Descent(noiseless)

(n, d, p, σ)	(1000,100,0.1,0.1)	(1000,100,0.2,0.1)	(1000,100,0.4,0.1)
Fixed	(0.03837,0.03855)	(0.04461,0.04458)	(0.06472,0.06469)
Reduced	(0.08775,0.08739)	(0.09009,0.09031)	(0.09249,0.09317)
EM-init	(0.09129,0.1139)	(0.04210,0.06023)	(0.04899,0.06161)
EM-rand	(0.1565,0.1791)	(0.2497,0.2875)	(0.7080,0.6076)
(n, d, p, σ)	(1000,100,0.1,0.3)	(1000,100,0.2,0.3)	(1000,100,0.4,0.3)
Fixed	(0.1267,0.1267)	(0.1536,0.1548)	(0.3721,0.3881)
Reduced	(0.2615,0.2623)	(0.2762,0.2760)	(0.3098,0.3118)
EM-init	(0.1930,0.2186)	(0.1948,0.2207)	(0.1853,0.2008)
EM-rand	(0.3327,0.3537)	(0.5340,0.5139)	(0.8175,0.7589)

(median,mean) over 100 trials(MLR) - Grad Descent(noisy)

Overview

Simulation Results - Low-dimensional case

$$y_i = x_i^T \beta_{z_i} + e_i,\ e_i \sim \mathcal{N}(0, \sigma^2)$$

(n, d, σ)	(30,3,0)	(60,3,0)	(90,3,0)
Fixed	(0.0620,0.246)	(5.207e-16,0.133)	(5.17e-16,0.100)
Reduced	(4.48e-16, 0.0242)	(4.80e-16,6.52e-16)	(4.25e-16,6.33e-16)
EM-init	(3.50e-16,0.0527)	(3.38e-16,0.00596)	(3.33e-16,0.0193)
EM-rand	(3.90e-16,0.167)	(3.68e-16,0.0754)	(3.4e-16,0.0648)
(n, d, s, σ)	(50,5,0)	(100,5,0)	(150,5,0)
Fixed	(0.340,0.347)	(7.15e-16,0.128)	(7.12e-16,0.0308)
Reduced	(7.91e-16,0.0258)	(6.76e-16,9.11e-16)	(6.74e-16,1.02e-15)
EM-init	(6.32e-16,0.0831)	(5.40e-16,0.0184)	(5.13e-16,0.0118)
EM-rand	(6.81e-16,0.190)	(5.43e-16,0.0348)	(5.26e-16,0.0386)
(n, d, s, σ)	(100,10,0)	(200,10,0)	(300,10,0)
Fixed	(2.79e-15,0.255)	(1.22e-15,0.0532)	(1.10e-15,0.0196)
Reduced	(1.218e-15,0.0138)	(1.13e-15,1.26e-15)	(1.12e-15,1.43e-15)
EM-init	(1.026e-15,0.0453)	(8.93e-16,0.0126)	(9.61e-16,1.09e-15)
EM-rand	(1.32e-15,0.234)	(9.06e-16,0.0555)	(9.74e-16,0.0299)

(median,mean) over 100 trials(MLR) - Full Correct

Simulation Results - Low-dimensional case

$$y_i = x_i^T \beta_{z_i} + e_i,\ e_i \sim \mathcal{N}(0, \sigma^2)$$

(n, d, σ)	(1000,100,0)	(500,100,0)	(2000,100,0)
Fixed	(0.7326,0.7225)	(0.7975,0.7929)	(3.977e-12,0.2660)
Reduced	(6.9652e-12,0.006843)	(0.6181,0.4955)	(3.568e-12,3.644e-12)
EM-init	(2.0896e-12,0.006398)	(0.6010,0.5323)	(1.810e-12,1.7538e-12)
EM-rand	(0.8138,0.6981)	(1.044,1.0391)	(1.998e-12,0.1267)

(median,mean) over 100 trials(MLR) - Grad Descent(noiseless)

(n, d, σ)	(1000,100,0.05)	(1000,100,0.1)	(1000,100,0.3)
Fixed	(0.7482,0.7141)	(0.7357,0.7140)	(0.7687,0.7720)
Reduced	(0.03332,0.03346)	(0.07626,0.07683)	(0.3488,0.4067)
EM-init	(0.02516,0.02520)	(0.05192,0.05894)	(0.1986,0.2239)
EM-rand	(0.8242,0.6745)	(0.8156,0.7188)	(0.8694,0.8076)

(median,mean) over 100 trials(MLR) - Grad Descent(noisy)

Overview

Simulation Results - High-dimensional case

(n, d, s, σ)	(200,1000,5,0)	(200,2000,5,0)	(200,2000,10,0)
Fixed	(0.6309,0.5142)	(0.5656,0.4758)	(0.8294,0.7917)
Reduced	(2.606e-12,0.2329)	(1.9967e-12,0.1903)	(0.7854,0.6780)
EM-rand	(0.7059,0.6198)	(0.8598,0.7714)	(0.9547,0.9660)

(median,mean) over 100 trials(MLR) - noiseless case

(n, d, s, σ)	(200,1000,5,0.05)	(200,1000,5,0.1)	(200,1000,5,0.3)
Fixed	(0.5302,0.4897)	(0.6341,0.5486)	(0.6648,0.6391)
Reduced	(0.02214,0.2270)	(0.05273,0.2802)	(0.3885,0.4501)
EM-rand	(0.6560,0.5450)	(0.7205,0.5980)	(0.7945,0.7487)

(median,mean) over 100 trials(MLR) - noisy case

Simulation Results - High-dimensional case

(n, d, s, σ)	(500,5000,10,0)	(1000,5000,20,0)
Fixed	(0.4928, 0.4285)	(0.7590,0.5964)
Reduced	(2.536e-13,0.02019)	(1.003e-13,0.02243)
EM-rand	(0.6856, 0.5507)	(0.3469,0.5763)

(median,mean) over 100 trials(MLR) - noiseless case

(n, d, s, σ)	(500,5000,10,0.05)	(500,5000,10,0.1)	(500,5000,10,0.3)
Fixed	(0.5699,0.4531)	(0.6228,0.5427)	(0.7077,0.6685)
Reduced	(0.01544,0.09110)	(0.03920,0.1083)	(0.3301,0.4112)
EM-rand	(0.8323,0.6422)	(0.7902,0.6005)	(0.9053,0.7963)

(median,mean) over 100 trials(MLR) - noisy case

References

Bhatia, Kush and Jain, Prateek and Kar, Purushottam(2015)
Advances in Neural Information Processing Systems
15, 721–729

The End