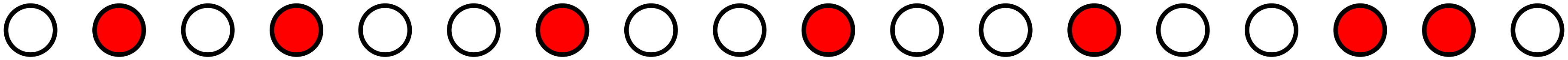
Learning Sparse Group Models Through Boolean Relaxation

Yijie Wang[1*], Yuan Zhou[2*], Xiaoqing Huang[3], Kun Huang[3], Jie Zhang[4], Jianzhu Ma[5]

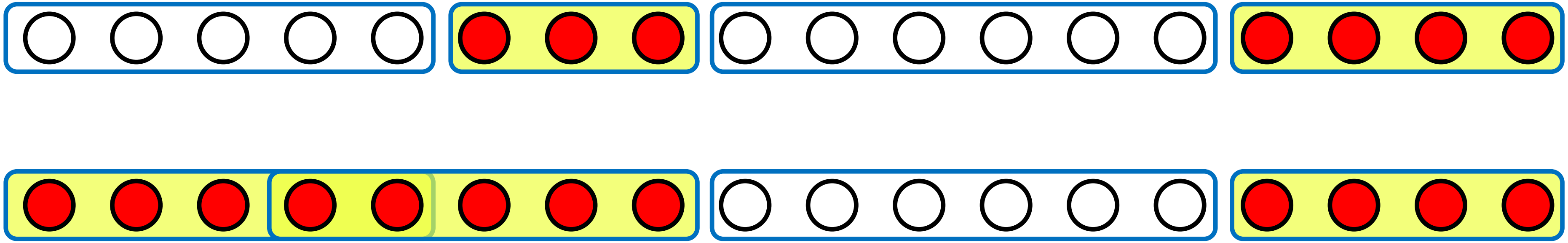
[1]Computer Science Department, Indiana University Bloomington
[2]Yau Mathematical Sciences Center and Department of Mathematical Sciences, Tsinghua University
[3]Department of Biostatistics & Health Data Science, Indiana University
[4]Department of Medical and Molecular Genetics, Indiana University
[5]Institute for AI Industry Research, Tsinghua University
[*]Equal contribution

Sparsity Learning

- Unstructured Sparsity



- Structured Sparsity



Sparse Group Models

- Exact formulation using constraints

$$P^* = \min_{w \in \Theta} \left\{ F(w) := \sum_{i=1}^{n} f(w^\top x_i; y_i) + \frac{1}{2}\rho \|w\|_2^2 \right\}$$

Sparse Group Models

- Exact formulation using constraints

Loss function

$$P^* = \min_{w \in \Theta} \left\{ F(w) := \sum_{i=1}^{n} \boxed{f(w^\top x_i; y_i)} + \frac{1}{2}\rho \|w\|_2^2 \right\}$$

Sparse Group Models

- Exact formulation using constraints

Regularization

$$P^* = \min_{w \in \Theta} \left\{ F(w) := \sum_{i=1}^{n} f(w^\top x_i; y_i) + \frac{1}{2}\rho \boxed{\|w\|_2^2} \right\}$$

Sparse Group Models

- Exact formulation using constraints

$$P^* = \min_{w \in \Theta} \left\{ F(w) := \sum_{i=1}^{n} f(w^\top x_i; y_i) + \frac{1}{2}\rho \|w\|_2^2 \right\}$$

Constraints

$$\Theta = \left\{ w \in \mathbb{R}^d \;\middle|\; \|w\|_0 \le k, \quad \sum_{j=1}^{b} \mathbf{1}\left[\left\|w_{g_j}\right\|_0 > 0\right] \le h \right\}$$

Sparse Group Models

- Exact formulation using constraints

$$P^* = \min_{w \in \Theta} \left\{ F(w) := \sum_{i=1}^{n} f(w^\top x_i; y_i) + \frac{1}{2}\rho \|w\|_2^2 \right\}$$

Constraints

$$\Theta = \left\{ w \in \mathbb{R}^d \;\middle|\; \|w\|_0 \leq k, \; \sum_{j=1}^{b} \mathbf{1}\left[\|w_{g_j}\|_0 > 0\right] \leq h \right\}$$

Constrain the # of selected individual features to be less than k

Constrain the # of selected groups of features to be less than h

Related Works

- Formulation using **regularization**
  - Structured sparsity-inducing norms (Friedman et al. (2010); Huang et al. (2011); Zhao et al. (2009); Simon et al. (2013); Tibshirani (1996); Bach et al. (2012); Kim & Xing (2012); Liu & Ye (2010); Rapaport et al. (2008); Zheng et al. (2018); Yuan et al. (2011); Jenatton et al. (2011))
  - Submodular set-functions (Bach (2010))
  - Convex relaxation of linear matrix inequalities and combinatorial penalties (El Halabi & Cevher (2015); Halabi et al. (2018))

- Formulation using **constraints**

Pilanci et al. (2015) --- Our special case	Our work
$P^* = \min_{\|w\|_0 \leq k} \left\{ F(w) := \sum_{i=1}^{n} f(w^\top x_i; y_i) + \frac{1}{2}\rho\|w\|_2^2 \right\}$	$P^* = \min_{w \in \Theta} \left\{ F(w) := \sum_{i=1}^{n} f(w^\top x_i; y_i) + \frac{1}{2}\rho\|w\|_2^2 \right\}$ $\Theta = \left\{ w \in \mathbb{R}^d \;\middle|\; \|w\|_0 \leq k, \quad \sum_{j=1}^{b} \mathbf{1}\left[\|w_{g_j}\|_0 > 0\right] \leq h \right\}$

Related Works

- Formulation using **regularization**
  - structured sparsity-inducing norms (Friedman et al. (2010); Huang et al. (2011); Zhao et al. (2009); Simon et al. (2013); Tibshirani (1996); Bach et al. (2012); Kim & Xing (2012); Liu & Ye (2010); Rapaport et al. (2008); Zheng et al. (2018); Yuan et al. (2011); Jenatton et al. (2011))
  - Submodular set-functions (Bach (2010))
  - Convex relaxation of linear matrix inequalities and combinatorial penalties (El Halabi & Cevher (2015); Halabi et al. (2018))

- Formulation using **constraints**

Pilanci et al. (2015) --- Our special case	Our work
$P^* = \min_{\|w\|_0 \le k} \left\{ F(w) := \sum_{i=1}^{n} f(w^\top x_i; y_i) + \frac{1}{2}\rho\|w\|_2^2 \right\}$	$P^* = \min_{w \in \Theta} \left\{ F(w) := \sum_{i=1}^{n} f(w^\top x_i; y_i) + \frac{1}{2}\rho\|w\|_2^2 \right\}$ $\Theta = \left\{ w \in \mathbb{R}^d \,\middle|\, \|w\|_0 \le k, \quad \sum_{j=1}^{b} \mathbf{1}\left[\|w_{g_j}\|_0 > 0\right] \le h \right\}$

Representation with Boolean constraints

- The original problem

$$P^* = \min_{w \in \Theta} \left\{ F(w) := \sum_{i=1}^{n} f(w^\top x_i; y_i) + \frac{1}{2}\rho \|w\|_2^2 \right\}$$

$$\Theta = \left\{ w \in \mathbb{R}^d \;\middle|\; \|w\|_0 \le k, \quad \sum_{j=1}^{b} \mathbf{1}\left[\left\| w_{g_j} \right\|_0 > 0 \right] \le h \right\}$$

- Exact representation with Boolean constraints (**Theorem 2.1**)

$$P^* = \min_{(u,z) \in \Gamma} \max_{v \in \mathbb{R}^n} \left\{ -\frac{1}{2\rho} v^\top X D(u) X^\top v - \sum_{i=1}^{n} f^*(v_i; y_i) \right\}$$

$$\Gamma = \left\{ (u, z) \;\middle|\; \sum_{i=1}^{d} u_i \le k, \quad \sum_{j=1}^{b} z_j \le h, \quad u_i \le z_j, \quad \forall i \in g_j, \quad u \in \{0,1\}^d, \quad z \in \{0,1\}^b \right\}$$

Representation with Boolean constraints

- The original problem

$$P^* = \min_{w \in \Theta} \left\{ F(w) := \sum_{i=1}^{n} f(w^\top x_i; y_i) + \frac{1}{2}\rho \|w\|_2^2 \right\}$$

$$\Theta = \left\{ w \in \mathbb{R}^d \,\middle|\, \|w\|_0 \le k, \quad \sum_{j=1}^{b} \mathbf{1}\left[\|w_{g_j}\|_0 > 0\right] \le h \right\}$$

- Exact representation with Boolean constraints (**Theorem 2.1**)

Legendre-Fenchel conjugate of f → $f^*(v_i; y_i)$

$$P^* = \min_{(u,z)\in\Gamma} \max_{v \in \mathbb{R}^n} \left\{ -\frac{1}{2\rho} v^\top X D(u) X^\top v - \sum_{i=1}^{n} f^*(v_i; y_i) \right\}$$

$$\Gamma = \left\{ (u, z) \,\middle|\, \sum_{i=1}^{d} u_i \le k, \quad \sum_{j=1}^{b} z_j \le h, \quad u_i \le z_j, \quad \forall i \in g_j, \quad u \in \{0,1\}^d, \quad z \in \{0,1\}^b \right\}$$

Representation with Boolean constraints

- The original problem

$$P^* = \min_{w \in \Theta} \left\{ F(w) := \sum_{i=1}^{n} f(w^\top x_i; y_i) + \frac{1}{2}\rho \|w\|_2^2 \right\}$$

$$\Theta = \left\{ w \in \mathbb{R}^d \,\middle|\, \|w\|_0 \le k, \quad \sum_{j=1}^{b} \mathbf{1}\left[\|w_{g_j}\|_0 > 0 \right] \le h \right\}$$

- Exact representation with Boolean constraints (**Theorem 2.1**)

$$P^* = \min_{(u,z) \in \Gamma} \max_{v \in \mathbb{R}^n} \left\{ -\frac{1}{2\rho} v^\top X D(u) X^\top v - \sum_{i=1}^{n} f^*(v_i; y_i) \right\}$$

$$\Gamma = \left\{ (u,z) \,\middle|\, \sum_{i=1}^{d} u_i \le k, \quad \sum_{j=1}^{b} z_j \le h, \quad u_i \le z_j, \quad \forall i \in g_j, \quad \boxed{u \in \{0,1\}^d}, \quad z \in \{0,1\}^b \right\}$$

u is a Boolean indicator for the supports of the individual features.

Representation with Boolean constraints

- The original problem

$$P^* = \min_{w \in \Theta} \left\{ F(w) := \sum_{i=1}^{n} f(w^\top x_i; y_i) + \frac{1}{2}\rho \|w\|_2^2 \right\}$$

$$\Theta = \left\{ w \in \mathbb{R}^d \;\middle|\; \|w\|_0 \leq k, \quad \sum_{j=1}^{b} \mathbf{1}\left[\left\|w_{g_j}\right\|_0 > 0 \right] \leq h \right\}$$

- Exact representation with Boolean constraints (**Theorem 2.1**)

$$P^* = \min_{(u,z) \in \Gamma} \max_{v \in \mathbb{R}^n} \left\{ -\frac{1}{2\rho} v^\top X D(u) X^\top v - \sum_{i=1}^{n} f^*(v_i; y_i) \right\}$$

$$\Gamma = \left\{ (u, z) \;\middle|\; \sum_{i=1}^{d} u_i \leq k, \quad \sum_{j=1}^{b} z_j \leq h, \quad u_i \leq z_j, \quad \forall i \in g_j, \quad u \in \{0,1\}^d, \quad \boxed{z \in \{0,1\}^b} \right\}$$

z is a Boolean indicator for the supports of the group features.

Boolean Relaxation

- Exact representation with Boolean constraints (**Theorem 2.1**)

$$P^* = \min_{(u,z)\in\Gamma} \max_{v\in\mathbb{R}^n} \left\{ -\frac{1}{2\rho} v^\top X D(u) X^\top v - \sum_{i=1}^{n} f^*(v_i; y_i) \right\}$$

$$\Gamma = \left\{ (u,z) \middle| \sum_{i=1}^{d} u_i \le k, \quad \sum_{j=1}^{b} z_j \le h, \quad u_i \le z_j, \quad \forall i \in g_j, \quad \boxed{u \in \{0,1\}^d, \quad z \in \{0,1\}^b} \right\}$$

- Boolean relaxation

$$P_{\mathrm{BR}} = \min_{(u,z)\in\Omega} \max_{v\in\mathbb{R}^n} \left\{ -\frac{1}{2\rho} v^\top X D(u) X^\top v - \sum_{i=1}^{n} f^*(v_i; y_i) \right\}$$

$$\Omega = \left\{ (u,z) \middle| \sum_{i=1}^{d} u_i \le k, \quad \sum_{j=1}^{b} z_j \le h, \quad u_i \le z_j, \quad \forall i \in g_j, \quad \boxed{u \in [0,1]^d, \quad z \in [0,1]^b} \right\}$$

The Tightness of the Boolean Relaxation

- In Theorem 2.2
  - For general loss function f
  - The sufficient and necessary condition when P_{BR} achieves the exact solution of P^*

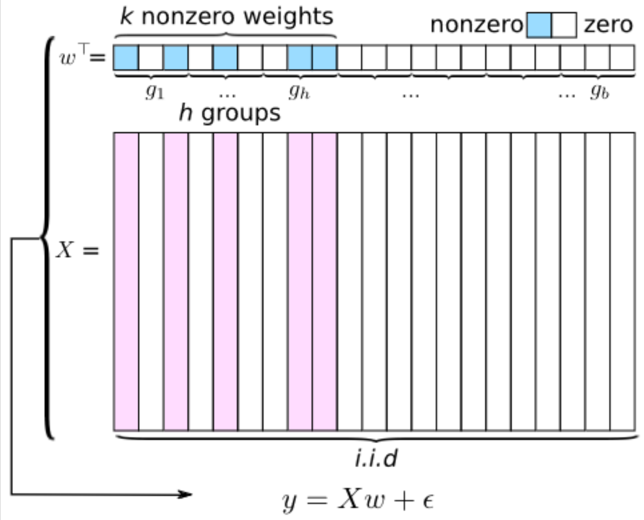
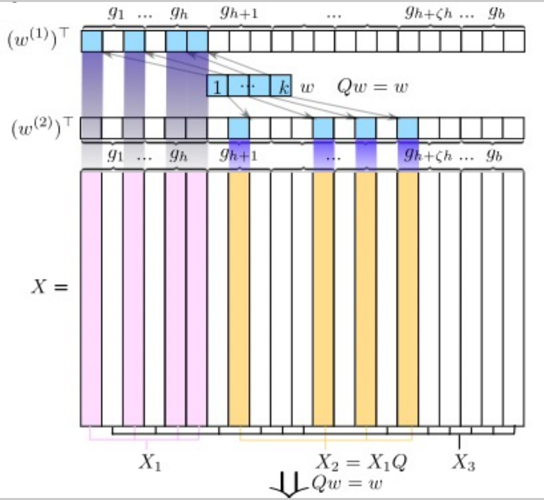
- In Corollary 2.3
  - For square loss $f(w^T x_i; y_i) = \frac{1}{2}(w^T x_i - y_i)^2$

$$L_{\mathrm{BR}} = \min_{(u,z)\in\Omega} \left\{ G(u) := y^\top \left(\frac{1}{\rho} X D(u) X^\top + I \right)^{-1} y \right\}$$

  - The sufficient and necessary condition when L_{BR} achieves the exact solution of L^*

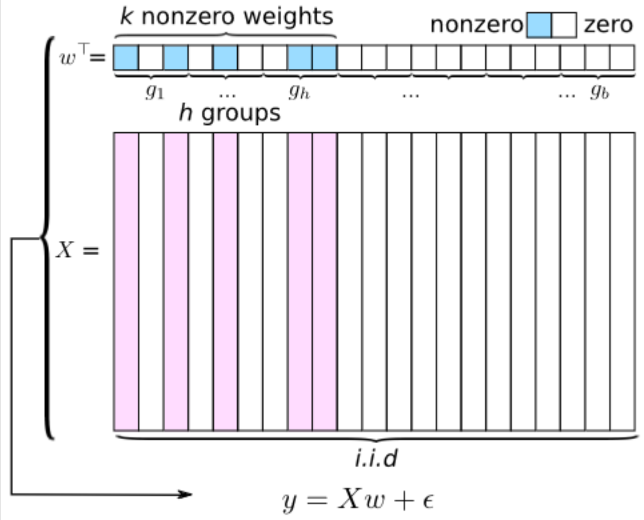
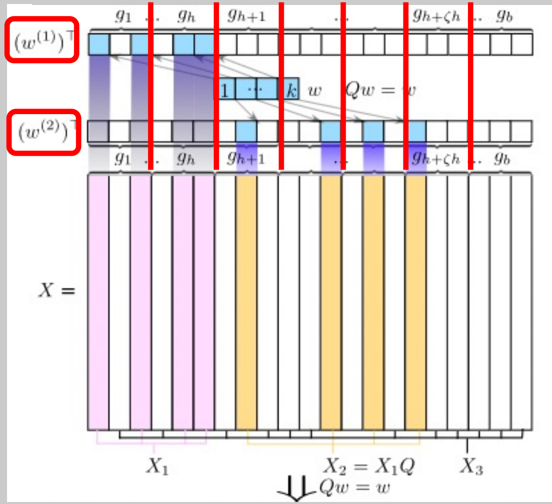
Theoretical Guarantees for Random Ensembles

- Apply **Corollary 2.3** to two **Random Ensembles**

Random Ensemble I (Simon et al. (2013); Friedman et al. (2010))	Random Ensemble II
	

Theoretical Guarantees for Random Ensembles

- Apply **Corollary 2.3** to two **Random Ensembles**

Random Ensemble I (Simon et al. (2013); Friedman et al. (2010))	Random Ensemble II
	

Theoretical Guarantees for Random Ensembles

- We prove our relaxed program
  - can achieves the exactness with high probability.
  - can achieve the nearly optimal sample complexity.

- Random Ensemble I

Theorem 3.1. *Consider the random instance described above with parameters* (n, d, k, γ, b, h) *and let* $y = Xw + \epsilon$ *be the observed response vector. Suppose that* $\gamma \geq 1$*. Let* $\rho = n^{1/2+\delta}$ $(\delta \in (0, 1/2))$*. With probability at least* $(1 - d\exp(-\Omega(n^{2\delta}/(\gamma^2 k))) - d\exp(-\Omega(n^{1-2\delta}))$*, the relaxed program* L_{BR} *admits the optimal solution* u^* *and* z^* *where* $u_i^* = \mathbf{1}[w_i \neq 0]$ *and* $z_j^* = \mathbf{1}[j \in \{1, 2, \ldots, h\}]$*.*

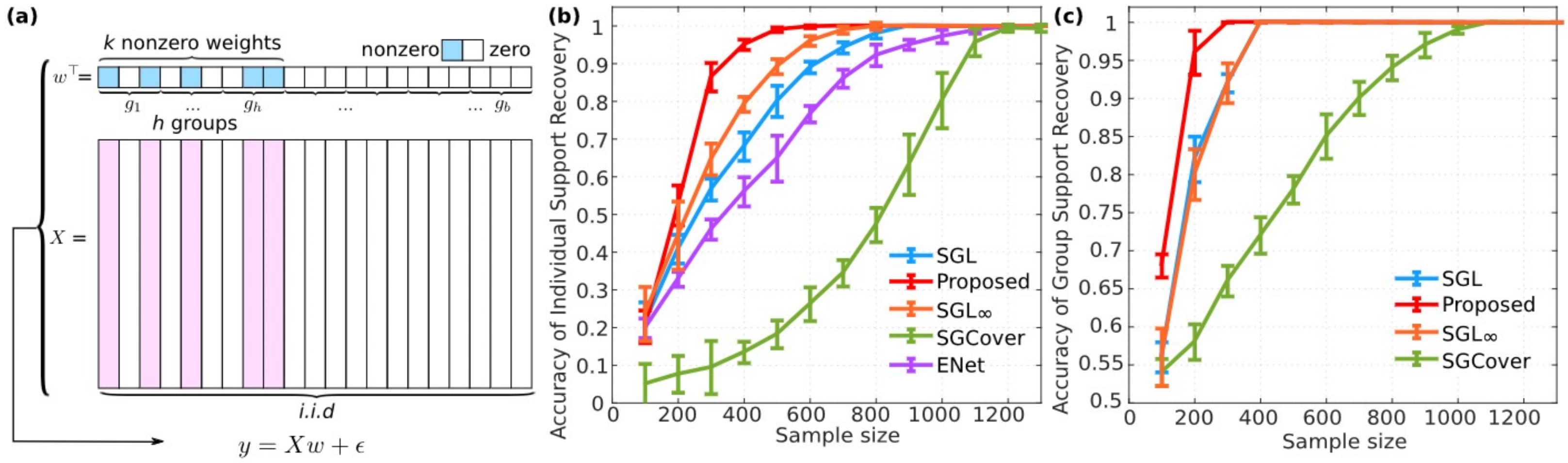
- Random Ensemble II

Theorem 3.2. *Let* $X = [X_1, X_2, X_3]$ *and* $y = Xw^{(1)} + \epsilon$ *be a random instance described above with parameters* $(n, d, k, \gamma, b, h, \zeta, w)$*. Suppose there exists* $\xi > 0$ *such that* $\xi \leq |w_i| \leq \zeta^{1/4}\xi$ *for all* $i \in \{1, 2, \ldots, k\}$*. Also suppose that* $\gamma \geq 1$*. Let* $\rho = n^{1/2+\delta}$ $(\delta \in (0, 1/2))$*. For large enough constant* ζ*, with probability at least* $(1 - d\exp(-\Omega(n^{2\delta}\xi^2/\gamma^2)) - d\exp(-\Omega(n^{1-2\delta})))$*, the relaxed program* L_{BR} *admits the optimal solution* u^* *and* z^* *where* $u_i^* = \mathbf{1}[w_i^{(1)} \neq 0]$ *and* $z_g^* = \mathbf{1}[\exists i \in g : w_i^{(1)} \neq 0]$*. Here, we use* g *to denote both the index of a group and the set of the features included in the group.*

Experimental Results

- Random Ensemble I

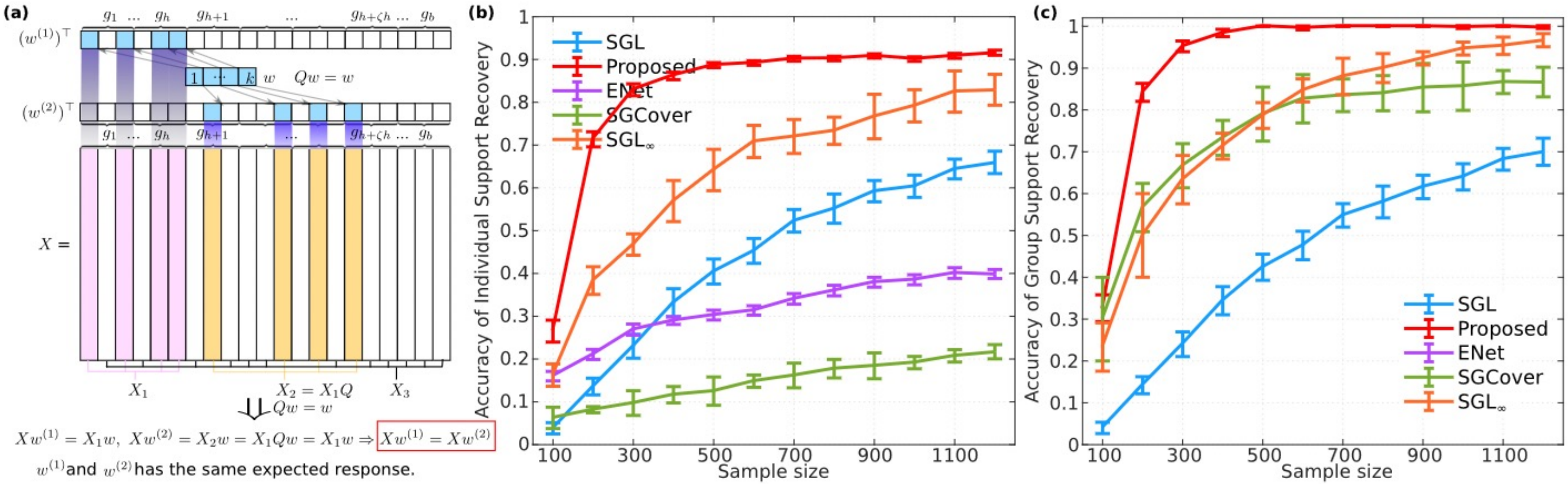
$$L_{\mathrm{BR}} = \min_{(u,z)\in\Omega} \left\{ G(u) := y^\top \left(\frac{1}{\rho} X D(u) X^\top + I \right)^{-1} y \right\}$$



Experimental Results

- Random Ensemble II

$$L_{\rm BR} = \min_{(u,z)\in\Omega} \left\{ G(u) := y^\top \left(\frac{1}{\rho} X D(u) X^\top + I \right)^{-1} y \right\}$$



Experimental Results

- Cancer drug response prediction
  - Drug: IMATNIB
  - Samples: IMATNIB response of 1,225 tumor samples
  - Features: 2,369 genes
  - Pathways: 207 gene groups

Table 1: Result comparison for IMATNIB.

Method	k(s.d.)	h(s.d.)	Out-of-sample MSE $\pm$95%CI
Proposed	46	7	32.6$\pm$ 2.2
SGL-Overlap	92 (5.4)	19 (0.5)	46.9$\pm$3.7
ENet	60 (8.2)	18 (2.3)	39.6 $\pm$ 4.2
SGCover	321 (10.5)	13 (1.7)	55.4 $\pm$ 6.9

Table S3: Pathways and genes identified by the proposed methods for IMATNIB.

Pathway	Genes	Reference
RHO GTPases Activate WASPs and WAVEs	ARPC1B WASF1 ARPC5 WASL CYFIP1 ACTG1 ACTR3	Gu et al. (2009); Huang et al. (2008); Chen et a
Regulation of PTEN gene transcription	LAMTOR3 LAMTOR4 SNAI1 RPTOR RRAGA RRAGB MBD3 RRAGD PHC3 GATAD2A RCOR1 MECOM CBX8 LAMTOR2	Nishioka et al. (2011); Peng et al. (2010); Huan
Signaling by PDGF	PDGFC COL4A3 COL6A2 COL6A3 COL9A3	Malavaki et al. (2013); Li et al. (2006); Heldin
Retinoid metabolism and transport	CLPS LRP8 APOC3 SDC4 LPL LRP10 LRP12 APOA2	Hoang et al. (2010)
TCF transactivating complex	RBBP5 KAT5 PYGO1 PYGO2 BCL9	Zhang et al. (2021); Coluccia et al. (2007); Co
Deactivation of the beta-catenin transactivating complex	RBBP5 SOX3 SRY PYGO1 PYGO2 CBY1 BCL9	Zhou et al. (2003); Leo et al. (2013)
RAS processing	ZDHHC9 GOLGA7 BCL2L1 ABHD17B	Chung et al. (2006); Braun & Shannon (2008)

Experimental Results

- Cancer drug response prediction
  - Drug: IMATNIB
  - Samples: IMATNIB response of 1,225 tumor samples
  - Features: 2,369 genes
  - Pathways: 207 gene groups

Table 1: Result comparison for IMATNIB.

Method	k(s.d.)	h(s.d.)	Out-of-sample MSE $\pm$95%CI
Proposed	46	7	32.6$\pm$ 2.2
SGL-Overlap	92 (5.4)	19 (0.5)	46.9$\pm$3.7
ENet	60 (8.2)	18 (2.3)	39.6 $\pm$ 4.2
SGCover	321 (10.5)	13 (1.7)	55.4 $\pm$ 6.9

Table S3: Pathways and genes identified by the proposed methods for IMATNIB.

Pathway	Genes	Reference
RHO GTPases Activate WASPs and WAVEs	ARPC1B WASF1 ARPC5 WASL CYFIP1 ACTG1 ACTR3	Gu et al. (2009); Huang et al. (2008); Chen et a
Regulation of PTEN gene transcription	LAMTOR3 LAMTOR4 SNAI1 RPTOR RRAGA RRAGB MBD3 RRAGD PHC3 GATAD2A RCOR1 MECOM CBX8 LAMTOR2	Nishioka et al. (2011); Peng et al. (2010); Huan
Signaling by PDGF	PDGFC COL4A3 COL6A2 COL6A3 COL9A3	Malavaki et al. (2013); Li et al. (2006); Heldin
Retinoid metabolism and transport	CLPS LRP8 APOC3 SDC4 LPL LRP10 LRP12 APOA2	Hoang et al. (2010)
TCF transactivating complex	RBBP5 KAT5 PYGO1 PYGO2 BCL9	Zhang et al. (2021); Coluccia et al. (2007); Co
Deactivation of the beta-catenin transactivating complex	RBBP5 SOX3 SRY PYGO1 PYGO2 CBY1 BCL9	Zhou et al. (2003); Leo et al. (2013)
RAS processing	ZDHHC9 GOLGA7 BCL2L1 ABHD17B	Chung et al. (2006); Braun & Shannon (2008)

Summary

- Novel framework for sparse group models.

- Theoretically for two random ensembles,
  - achieve the exactness with high probability.
  - achieve nearly optimal sample complexity.

- Empirically,
  - outperforms the state-of-the-art methods when the sample size is small.

Thank you!!!

Rounding Scheme

- Recover Boolean solution ($u \in \{0,1\}^d, z \in \{0,1\}^b$) from ($\bar{u} \in [0,1]^d, \bar{z} \in [0,1]^b$)

- Rounding Algorithm
  - Generate feasible Boolean solution $(\tilde{u}, \tilde{z})$
    - For group j, $\Pr[z_j = 1] = \bar{z}_j \quad \text{and} \quad \Pr[z_j = 0] = 1 - \bar{z}_j.$
    - For feature i in group j,

$$\Pr[u_i = 1] = \frac{\bar{u}_i}{\bar{z}_j} \quad \text{and} \quad \Pr[u_i = 0] = 1 - \frac{\bar{u}_i}{\bar{z}_j}.$$

  - Find the best solution

$$w := \arg\min_{w \in \mathbb{R}^d} F(D(\tilde{u})w)$$