Logic Rules for Sparse Learning

Anna Deza[1] Andrés Gómez[2] Alper Atamtürk[1]

[1]IEOR, University of California, Berkeley, CA, USA [2]ISE, University of Southern California, Los Angeles, CA, USA

Berkeley UNIVERSITY OF CALIFORNIA

Sparse Learning

Sparse Learning: aim to build models that retain *only key most informative features*, discarding rest

- Important for interpretability and generalization performance
- Useful in settings where number of features (p) >> number of samples

Example: Using genetic data to estimate health risk. Practitioners desire interpretability.

Dense Solution: Hard to interpret ✗

Sparse Solution: Interpretable ✓

We build models by finding model coefficients $\beta \in \mathbb{R}^p$ that minimize loss function $\mathcal{L}(\beta)$.

Sparsity of model can be represented by $\|\boldsymbol{\beta}\|_0$.

We consider regularized and cardinality constrained sparse learning

$$\min_{\beta \in \mathbb{R}^p} \mathcal{L}(\beta) + \lambda\|\beta\|_2^2 + \mu\|\boldsymbol{\beta}\|_0 \qquad \text{(REG)}$$

$$\min_{\beta \in \mathbb{R}^p} \mathcal{L}(\beta) + \lambda\|\beta\|_2^2 \text{ s.t. } \|\boldsymbol{\beta}\|_0 \leq k. \qquad \text{(CARD)}$$

Key challenge: exact sparsity penalty makes this **NP-hard**.

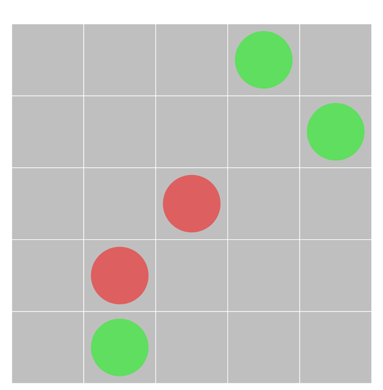
- Recent interest in using *mixed-integer optimization* techniques to solve to optimality

Background: Screening Rules

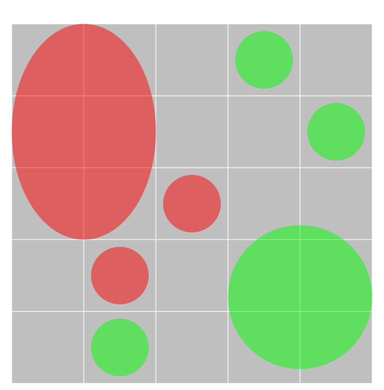
Safe screening [1]: Techniques that eliminate features guaranteed to not be in the optimal solution, *reducing the dimension* of the problem before the full optimization step, *improving solution times*.
→ These methods use lower and upper bounds on optimal values to fix features
→ Effectiveness *degrades* when *relaxation gaps are large*.

Generalizing Screening: Logic Rules

- Idea: Instead of considering inclusion/exclusion of a single feature independently from the rest, consider the **logical relationships** between groups of features as well.



Screening Rules only



Screening + **Logic** Rules

Logic rules consider

- inclusion of a group of features,
- exclusion of a group of features,
- ranking of pairs of features

generalizing screening rules and are able to handle *larger gaps*.

Logic Rules: A Preprocesing Step

Stage 0: Formulate sparse learning as a *mixed integer program* by introducing $z_i \in \{0, 1\}$ to model

$$z_i = 0 \Rightarrow \beta_i = 0.$$

Applying Logic Rules is a two stage process:

Stage 1

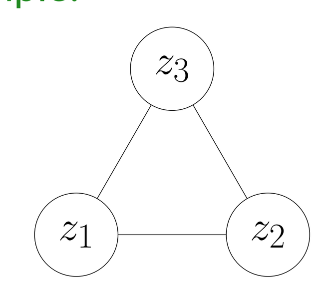
Find logical relationships between *pairs* of features, equivalent to exclusivity constraints.
Example:

$$z_1 + z_2 \leq 1$$
$$z_1 + z_3 \leq 1$$
$$z_2 + z_3 \leq 1$$
$$z \in [0, 1]^3$$

Key: Our method only uses the solution to the **relaxation** of the problem

Stage 2

Construct stronger inequalities implied by collection of constraints, equivalent to finding maximal cliques in a conflict graph.
Example:



$$z_1 + z_2 + z_3 \leq 1$$
$$z \in [0, 1]^3$$

Key: We exploit *special structure* to **efficiently** find all maximal cliques in $\mathcal{O}(p \log p)$

Key Takeaway

We propose a **general preprocessing** framework that *generates inequalities* that can be leveraged by mixed integer optimization solvers to speed up sparse learning computation.

1. Inequalities identified are of a *special structure* that solvers leverage to **improve computation**
2. Proposed method is **efficient** due to the exploitation of *an underlying structure* (chordality) in the conflict graph generated by the inequalities
3. Helps where screening rules are unsuccessful by **remaining effective** when relaxation **gaps are large**, while requiring *negligible additional computation*

Applications

Many applications: Encapsulates many learning models such as sparse regression, binary classification, multi-class logistic models.

Healthcare, Genomics, Finance, Image Sensing, Natural Language Processing, Climate Forecasting

Proposition: *Safe Logic Rules for Regularized Learning*

Let ζ_R be the relaxation objective value of (REG), α a value computed from its optimal solution, and ζ_u an upper bound. Then any optimal solution z to (REG) satisfies the following rule on the right given the corresponding condition holds.

Condition	Logic Rule
$\zeta_R + \alpha_i + \alpha_j > \zeta_u$	$z_i + z_j \leq 1$
$\zeta_R - \alpha_i - \alpha_j > \zeta_u$	$z_i + z_j \geq 1$
$\zeta_R + \alpha_i - \alpha_j > \zeta_u$	$z_i \leq z_j$
$\zeta_R - \alpha_i + \alpha_j > \zeta_u$	$z_i \geq z_j$

Key Computational Results

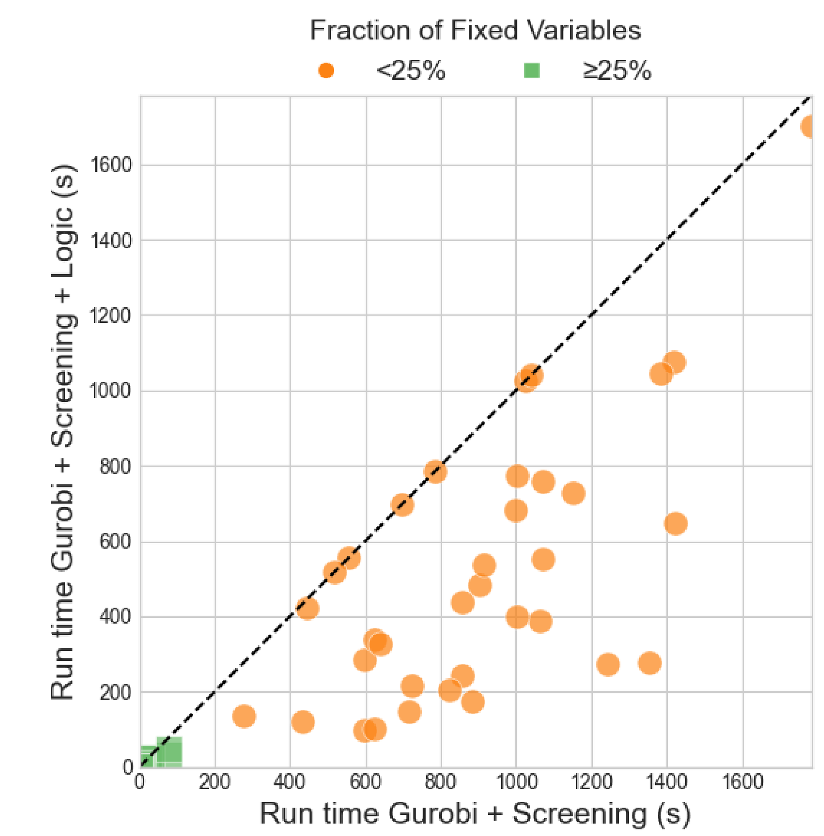
- On synthetic data, observe 50% **reduction** in branch-and-bound **nodes** when using logic rules, *double* what is observed for using screening alone.
- Observe negligible improvement in runtimes when screening is highly effective ($545\times \to 585\times$), but when screening fails observe $3\times$ **runtime speedup** over Gurobi alone.
- On real data experiments, logic rules provide $3\times$ **speedup** over screening rules.
- For real data instances which do not terminate within an hour time-limit logic rules give **better optimality gaps** of 13% vs 52%, coming from better solutions found within the limit.
- Advantage over screening rules is in instances that are noisier and have weaker regularizing, oweing to larger relaxation gaps.

Computational Results Details

We generate synthetic data with $p = 1000$ and 100 observations. We compare runtimes for solving (CARD) using screening rules alone and screening in conjunction with logic rules.

We vary noise levels (SNR) and regularization strength (λ) to show the regions in which logic rules add additional computation gains.

Real data experiments are done on genomic data with $p = 4,088$ and 71 observations.



SNR	λ	RGap %	Gurobi Runtime (s)	Gurobi + Screen Runime (s)	Gurobi + Screen + Logic Runtime (s)
0.05	1/10	47.4	1,572	1,574	981
	1/8	31.6	1,083	1,083	581
	1/4	7.5	761	4.1	3.8
	1/2	1.6	439	0.7	0.7
1.0	1/10	39.8	728	733	482
	1/8	28.5	715	505	266
	1/4	7.5	646	7.8	6.4
	1/2	1.5	386	0.6	0.6
6.0	1/10	25.5	684	276	57
	1/8	20.3	755	163	40
	1/4	6.0	605	1.0	1.0
	1/2	1.4	527	0.5	0.5
Average		18.2	741	362	202

[1] A. Atamtürk and A. Gómez, "Safe screening rules for ℓ_0-regression from perspective relaxations," PMLR, 2020.