Experimental Design on a Budget for Sparse Linear Models

Sanjan Gupta
(Reed Group)

6th Oct, 2017
Systems Seminar

Regression terminology

- Rent = f(Area, Locality ...)

Area	Bus stop	Amenities	Locality	Rent
550	0.1	1	1	610
600	0.09	1	1	660
850	0.12	4	1	1030
800	0.2	3	3	980
950	0.18	5	3	1210

Data point, x

Features

Target variable, y

Regularized linear regression

- To build a linear model:

$$y_i = x_i^T \beta + \epsilon$$

where $x_i, \beta \in \mathbb{R}^p, y_i \in \mathbb{R}$ and $\epsilon \sim \mathcal{N}(0,1)$

- Regression setting:

$$\beta^* := \arg\min_{\beta} \frac{1}{2} ||X\beta - y||_u^v + \epsilon g(\beta)$$

Least squares regn loss function for v = u = 2.

Penalty function

Choices for penalty function

- Ridge regression:

$$g(\beta) = \beta^T M \beta \qquad \text{for some } M \succ 0$$

- closed form solution exists

$$\beta^* = \left(X^T X + \epsilon M\right)^{-1} X^T y$$

- LASSO (Least Absolute Shrinkage & Selection Operator):

$$g(\beta) = ||\beta||_1$$

Choices for penalty function

- LASSO often leads to unstable solutions

β_1	β_2	Ridge penalty	LASSO penalty
0	1	1	1
0.2	0.8	0.68	-
0.5	0.5	0.5	-
0.8	0.2	0.68	-
1	0	1	1

- Elastic net regression:

$$g(\beta) = \lambda||\beta||_1 + (1-\lambda)||\beta||_2$$

Objective

- Given a fixed budget, B identify a set of x_i's for which corresponding y_i's would be obtained experimentally.

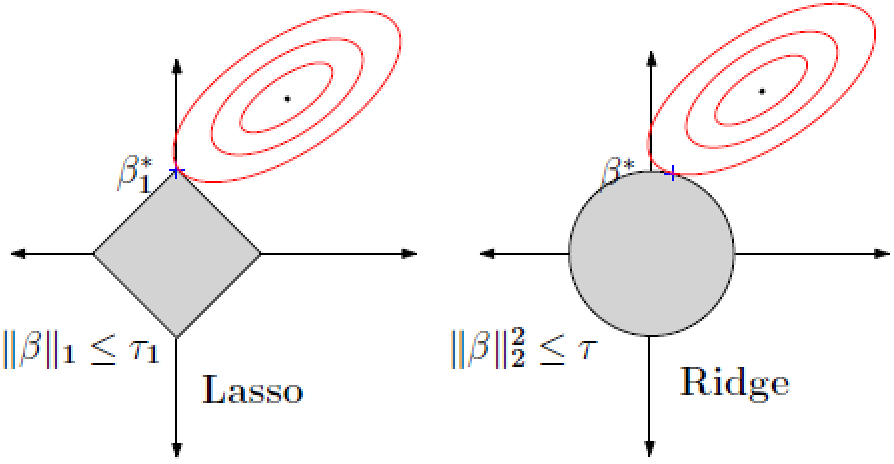
ED-S: Spectral Experimental Design

$$\beta^* = \arg\min_{\beta} \frac{1}{2}||X\beta - y||_2^2 + \lambda||\beta||_2^2 \quad \text{(RIDGE)}$$

$$\equiv \arg\min_{\beta} \frac{1}{2}||X\beta - y||_2^2 \ \text{ s.t. } \ ||\beta||_2^2 \leq \tau$$

$$\beta_1^* = \arg\min_{\beta} \frac{1}{2}||X\beta - y||_2^2 + \lambda||\beta||_1 \quad \text{(LASSO)}$$

$$\equiv \arg\min_{\beta} \frac{1}{2}||X\beta - y||_2^2 \ \text{ s.t. } \ ||\beta||_1 \leq \tau_1$$



- Typical Experimental Design (ED) formulation:

$$S^* := \arg \max_{|S| \leq B} f\left(\sum_{i \in \mathcal{S}} x_i x_i^T + \epsilon I\right)$$

i.e. identify S containing the set of selected subjects for a given budget B

$$S^* = \arg \max_{\mu \in \{0,1\}^n} f\left(\sum_{i=1}^{n} \mu_i x_i x_i^T + \epsilon I\right) \text{s.t.} \quad \mathbf{1}^T \mu \leq B$$

where $\mathbf{1} \in \mathbb{R}^n$

[Note: A popular choice for f(.) = log det (.)]

- Typical Experimental Design (ED) formulation:

$$S^* := \arg \max_{|S| \le B} f\left(\sum_{i \in \mathcal{S}} x_i x_i^T + \epsilon I\right)$$

i.e. identify S containing the set of selected subjects for a given budget B

$$S^* = \arg \max_{\mu \in \{0,1\}^n} f\left(\sum_{i=1}^{n} \mu_i x_i x_i^T + \epsilon I\right) \text{s.t.} \quad \mathbf{1}^T \mu \le B$$

where $\mathbf{1} \in \mathbb{R}^n$

[Note: A popular choice for f(.) = log det (.)]

- Spectral ED formulation:

$$\min_{\mu, u} \log\det\left(\sum_{i=1}^{n} \mu_i x_i x_i^T + \epsilon I\right)^{-1} + \lambda ||\gamma - u||_2^2$$

$$\text{s.t. } 0 \le \mu \le 1, \mathbf{1}^T \mu \le B,$$

$$u \in \arg\max_{v} \left\{ v^T \left(\sum_{i=1}^{n} \mu_i x_i x_i^T + \epsilon I\right) v \text{ s.t. } v^T v = 1 \right\}$$

- For a symmetric positive definite matrix, M the largest eigenvector is given by:

$$\arg\max_{u:||u||_2^2=1} u^T M u = \arg\min_{u:||u||_2^2=1} u^T M^{-1} u$$

- ED-S upon reformulation:

S =

$$\begin{aligned} &\min_{\mu,u} \log\det\left(\sum_{i=1}^{n}\mu_i x_i x_i^T + \epsilon I\right)^{-1} \\ &\quad + \lambda||\gamma - u||_2^2 + u^T\left(\sum_{i=1}^{n}\mu_i x_i x_i^T + \epsilon I\right)^{-1} u \\ &\text{s.t. } 0 \leq \mu \leq 1, \quad \mathbf{1}^T\mu \leq B, \quad ||u||_2^2 = 1 \end{aligned}$$

Non-convex problem

- ED-S formulation:

S =

$$\min_{\mu,u} \log\det\left(\sum_{i=1}^{n}\mu_i x_i x_i^T + \epsilon I\right)^{-1}$$

$$+ \lambda||\gamma - u||_2^2 + u^T\left(\sum_{i=1}^{n}\mu_i x_i x_i^T + \epsilon I\right)^{-1} u$$

$$\text{s.t. } 0 \le \mu \le 1, \quad \mathbf{1}^T\mu \le B, \quad ||u||_2^2 = 1$$

Non-convex problem

- Define sub-problems:
  - Fix μ to get S_O
  - Fix u to get S_μ

Algorithm 1 Alternating Minimization Algorithm

Pick arbitrary starting point μ, initialize u such that $||u||_2 = 1$.
for $t = 1, 2, \cdots, T$ **do**
  Update $\mu \leftarrow \arg\min S_\mu$
  Update $u \leftarrow \arg\min S_O$
end for

Some extensions:

- If Hessian matrix is large, use top k eigenvectors:

$$S = \min_{\mu,u} \log\det\left(\sum_{i=1}^{n} \mu_i x_i x_i^T + \epsilon I\right)^{-1} + \lambda||\gamma - u||_2^2 + u^T\left(\sum_{i=1}^{n} \mu_i x_i x_i^T + \epsilon I\right)^{-1} u$$

$$\text{s.t. } 0 \leq \mu \leq 1, \quad \mathbf{1}^T\mu \leq B, \quad ||u||_2^2 = 1$$

$$S_O = \min_{u:||u||_2^2=1} \lambda||\gamma - u||_2^2 + u^T M u$$

where M = $\left(\sum_{i=1}^{k} \mu_i x_i x_i^T + \epsilon I\right)^{-1}$.

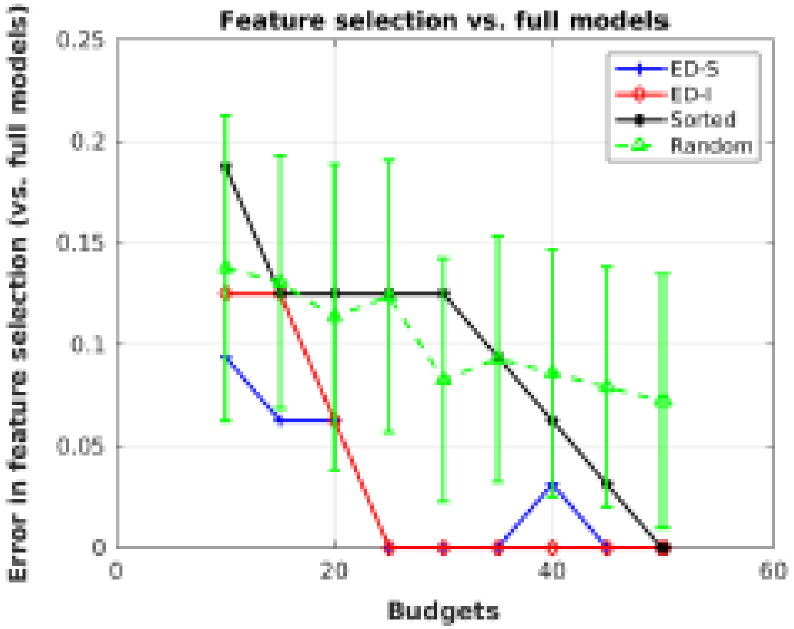
$$\min_{U\in\mathbb{R}^{p\times k}} \sum_{j=1}^{k} ||\gamma_j - u_j||_2^2 + tr\left(U^T M U\right) \quad \text{s.t. } U^T U = I$$

- Number of decision variables = **pk + n** where p = # of features

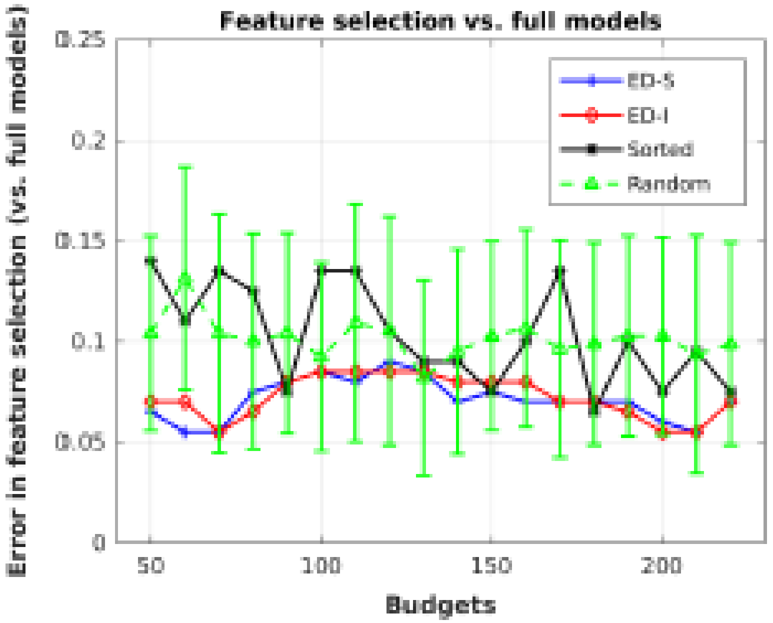
  Propose ED-I strategy for p >> n

Comparison to baseline designs

(a)



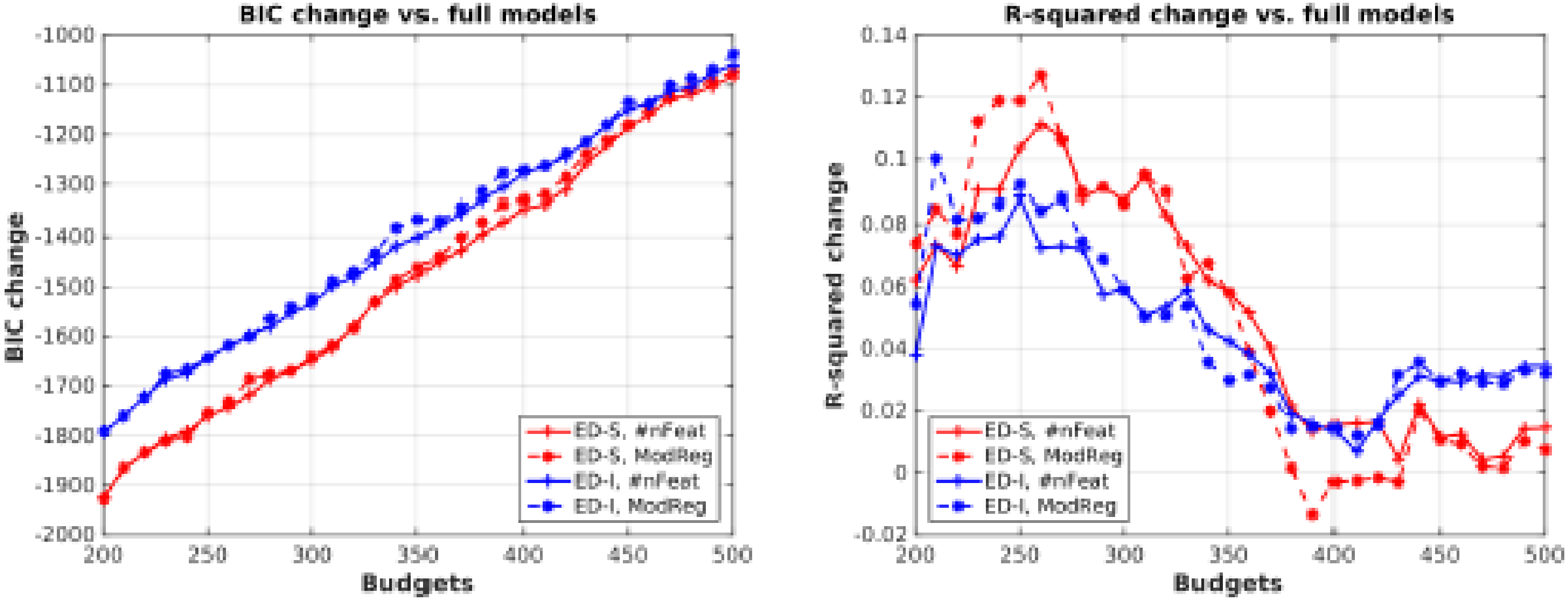
(b)



Error in consistent selection of correct features (derived from full model) for *prostate* and *lars* datasets having 8 and 10 features resp.

Neuro dataset

- 118 Region-of-interest based features (PET scans); ~1000 subjects



- Potential cost savings in longitudinal clinical trials

Reference

Experimental Design on a Budget for Sparse Linear Models and Applications

Sathya N. Ravi[†] RAVI5@WISC.EDU
Vamsi K. Ithapu[†] ITHAPU@WISC.EDU
Sterling C. Johnson[§†] SCJ@MEDICINE.WISC.EDU
Vikas Singh[†] VSINGH@BIOSTAT.WISC.EDU
[†]University of Wisconsin Madison [§†]William S. Middleton Memorial Veterans Hospital

Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA, 2016. JMLR: W&CP volume 48.

Thank
You