High-dimensional Cost-constrained Regression via Non-convex Optimization

Guan Yu

June 5th, 2018

Department of Biostatistics
The State University of New York at Buffalo

This is joint work with Dr. Yufeng Liu and Dr. Haofa Fu

Outline

- Research Problem
- Method
- Theoretical Properties
- Simulation Study
- Application to a diabetes study

High Dimensional Cost-constrained Regression

Consider the following linear regression model

$$y = x^T \beta^0 + \epsilon,$$

where

$$x = (x_1, x_2, \ldots, x_p)^T,\ \mathrm{E}(x) = 0, \mathrm{Cov}(x) = \mathbf{\Sigma},\ \mathrm{E}(\epsilon) = 0, \mathrm{Var}(\epsilon) = \sigma^2.$$

In practice, we need to spend money on collecting data. Suppose we need to spend c_j dollars on collecting the value of the j-th predictor x_j and our budget is C dollars. Assume that C is small and therefore our proposed model can only use a few predictors.

Question:

How to find a predictive linear model that satisfies the budget constraint and has good prediction performance?

High Dimensional Cost-constrained Regression

The regression coefficient vector of the best cost-constrained regression model is

$$
\begin{aligned}
\beta^* &\in \arg\min_{\beta} \mathrm{E}[(y - \sum_{j=1}^{p} x_j \beta_j)^2] \text{ subject to } \sum_{j=1}^{p} c_j \mathcal{I}(\beta_j) \le C \\
&= \arg\min_{\beta} \mathrm{E}[(\sum_{j=1}^{p} x_j \beta_j^0 + \epsilon - \sum_{j=1}^{p} x_j \beta_j)^2] \text{ subject to } \sum_{j=1}^{p} c_j \mathcal{I}(\beta_j) \le C \\
&= \arg\min_{\beta} (\beta - \beta^0)^T \mathbf{\Sigma} (\beta - \beta^0) \text{ subject to } \sum_{j=1}^{p} c_j \mathcal{I}(\beta_j) \le C,
\end{aligned}
$$

where $\mathcal{I}(\beta_j)$ is an indicator function which equals to 1 if $\beta_j \neq 0$ and 0 otherwise.

High Dimensional Cost-constrained Regression

Given the training data $\{Y, \mathbf{X}\}$, it is natural to estimate β^* by solving the following sample-average approximation (SAA) problem

$$\min_{\beta} \frac{1}{2n} \|Y - \mathbf{X}\beta\|_2^2 \text{ subject to } \sum_{j=1}^{p} c_j \mathcal{I}(\beta_j) \leq C.$$

- This problem can be viewed as a generalized best subset selection problem (the case with $c_1 = c_2 = \cdots = c_p$).
- The constraint $\sum_{j=1}^{p} c_j \mathcal{I}(\beta_j) \leq C$ makes this problem NP-hard.
- Even for the best subset selection problem, in order to find the global solution, most state-of-the-art algorithms do not scale to problems with more than 30 variables (Bertsimas et al. (2016)).

The difference between β^* and β^0

Since we aim to find the best cost-constrained regression model, the parameter of interest is β^* rather than β^0 in the true linear model.

$$\beta^* = \arg\min_{\beta} (\beta - \beta^0)^T \mathbf{\Sigma} (\beta - \beta^0) \text{ subject to } \sum_{j=1}^{p} c_j \mathcal{I}(\beta_j) \leq C,$$

Denote $S = \{j : \beta_j^0 \neq 0\}$.

- If $\sum_{j=1}^{p} c_j \mathcal{I}(\beta_j^0) = \sum_{j \in S} c_j \leq C$, then we know that β^0 is a feasible solution and $\beta^* = \beta^0$.
- If $\sum_{j \in S} c_j > C$ and $\mathrm{Cov}(x_S, x_{S^c}) = 0$, we can prove that $\beta^*_{S^c} = 0$ and

$$\beta^*_S = \arg\min_{\beta} (\beta_S - \beta^0_S)^T \mathbf{\Sigma}_{SS} (\beta_S - \beta^0_S)$$
$$\text{subject to } \sum_{j \in S} c_j \mathcal{I}(\beta_j) \leq C.$$

Could we use a two-step method?

For the above two cases, we have $\beta^*_{S^c} = \beta^0_{S^c} = 0$.

Therefore, if β^0 is sparse, we can use the following two-step method:

Step 1: Perform a screening to find a subset $\hat{S}$ which contains the true subset S with high probability and let $\hat{\beta}^*_{\hat{S}^c} = 0$;

Step 2: Solve the following optimization problem

$$\hat{\beta}^*_{\hat{S}} = \arg\min_{\beta_{\hat{S}}} \frac{1}{2n} \|Y - \mathbf{X}_{\hat{S}} \beta_{\hat{S}}\|_2^2$$

$$\text{subject to } \sum_{j \in \hat{S}} c_j \mathcal{I}(\beta_j) \leq C.$$

Could we use a two-step method?

However, in many cases, $\beta^*_{S^c}$ is not equal to 0 and therefore, we can not use the two-step method.

Let $p = 3, \beta = (\beta_1, \beta_2, \beta_3)^T$ and $\beta^0 = (t_0, t_0, 0)$ where $t_0 > 0$. Assume that the covariance matrix

$$\Sigma = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}$$

If $-\frac{1}{2} < \rho < -\frac{1}{3}$, we can show that $\beta^* = (0, 0, 2\rho t_0)$.

Therefore, $\{j : \beta^0_j \neq 0\} \cap \{j : \beta^*_j \neq 0\}$ is an empty set!

Method: Orthogonal Case

Assume that $n > p$ and $\mathbf{X}^T\mathbf{X} = n\mathbf{I}_n$. Under these assumptions, we can show that

$$\begin{aligned}\frac{1}{2n}\|Y - \mathbf{X}\beta\|_2^2 &= \frac{1}{2n}\|Y - \mathbf{X}\tilde{\beta}\|_2^2 + \frac{1}{2n}\|\mathbf{X}\beta - \mathbf{X}\tilde{\beta}\|_2^2 \\ &= \frac{1}{2n}\|Y - \mathbf{X}\tilde{\beta}\|_2^2 + \frac{1}{2}\|\beta - \tilde{\beta}\|_2^2,\end{aligned}$$

where $\tilde{\beta} = \mathbf{X}^T Y/n$ is the least squares estimate of the true regression coefficient β^0.

Therefore, the original problem is equivalent to the following optimization problem

$$\min_{\beta} \|\beta - \tilde{\beta}\|_2^2 \quad \text{subject to } \sum_{j=1}^{p} c_j \mathcal{I}(\beta_j) \leq C. \tag{1}$$

How to solve problem (1)?

If $\hat{\beta}$ is an optimal solution to the following problem

$$\min_{\beta} \|\beta - a\|_2^2 \quad \text{subject to} \ \sum_{j=1}^{p} c_j \mathcal{I}(\beta_j) \leq C,$$

then $\hat{\beta} = a \circ \hat{Z}$ where $\circ$ denotes the entrywise product of two vectors, and $\hat{Z} = (\hat{z}_1, \hat{z}_2, \ldots, \hat{z}_p)$ is the solution to the following 0-1 knapsack problem

$$\max_{z_1, z_2, \ldots, z_p} \sum_{j=1}^{p} a_j^2 z_j \quad \text{subject to} \ \sum_{j=1}^{p} c_j z_j \leq C, \ \text{and} \ z_1, z_2, \ldots, z_p \in \{0, 1\}.$$

Therefore, to solve problem (1), we only need to solve a 0-1 knapsack problem.

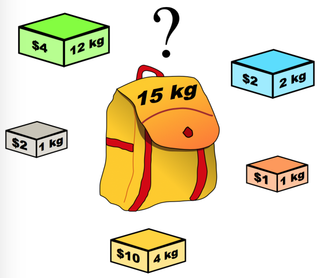
0-1 knapsack problem and dynamic programming

0-1 Knapsack Problem

$$\max_{z_1, z_2, \ldots, z_p} \sum_{j=1}^{p} v_j z_j$$

$$\text{subject to } \sum_{j=1}^{p} c_j z_j \leq C,$$

$$\text{and } z_1, z_2, \ldots, z_p \in \{0, 1\}.$$



Dynamic Programming

Denote $f[j, w]$ be the maximum value that can be attained with weight less than or equal to w using items up to j. We can define $f[j, w]$ recursively as follows:

- $f[0, w] = 0$;
- $f[j, w] = f[j-1, w]$ if $c_j > w$;
- $f[j, w] = \max(f[j-1, w], f[j-1, w-c_j] + v_j)$ if $c_j \leq w$.

The solution can then be found by calculating $f[p, C]$. The 0-1 knapsack problem can be solved in pseudo-polynomial time ($O(pC)$) using dynamic programming.

Special case: $c_1 = c_2 = \cdots = c_p = 1$

(**Bertsimas et al. (2016), Annals of Statistics**)

If $\hat{\beta}$ is an optimal solution to the following problem:

$$\hat{\beta} \in \arg\min_{\|\beta\|_0 \leq C} \|\beta - \tilde{\beta}\|_2^2,$$

then it can be computed as follows: if $|\tilde{\beta}_{(1)}| \geq |\tilde{\beta}_{(2)}| \geq \cdots \geq |\tilde{\beta}_{(p)}|$ denote the ordered values of the absolute values of the vector $\tilde{\beta}$, then

$$\hat{\beta}_j = \begin{cases} \tilde{\beta}_j, \text{ if } j \in \{(1), (2), \ldots, (C)\}, \\ 0, \text{ otherwise.} \end{cases}$$

Method: General Case

To solve the general high-dimensional cost-constrained regression problem, we use projected gradient descent methods.

- Denote $g(\beta) = \frac{1}{2n}\|Y - \mathbf{X}\beta\|_2^2$ and $L = \lambda_{max}(\frac{\mathbf{X}^T\mathbf{X}}{n})$, we have

$$g(\eta) \leq Q_L(\eta, \beta) := g(\beta) + \frac{L}{2}\|\eta - \beta\|_2^2 + \langle \nabla g(\beta), \eta - \beta \rangle$$

  for all β, η with equality holding at $\beta = \eta$.

- Given a current solution $\beta^{(m)}$, we upper bound the function $g(\eta)$ by the function $Q_L(\eta, \beta^{(m)})$, and update the solution by

$$\beta^{(m+1)} = \arg\min_{\eta} Q_L(\eta, \beta^{(m)}) \text{ subject to } \sum_{j=1}^{p} c_j \mathcal{I}(\eta_j) \leq C.$$

Method: General Case

We can show that

$$\begin{aligned}\beta^{(m+1)} &= \arg\min_{\eta} Q_L(\eta, \beta^{(m)}) \text{ subject to } \sum_{j=1}^{p} c_j \mathcal{I}(\eta_j) \le C \\ &= \arg\min_{\eta} \|\eta - (\beta^{(m)} - \frac{1}{L}\nabla g(\beta^{(m)}))\|_2^2 \\ &\quad \text{subject to } \sum_{j=1}^{p} c_j \mathcal{I}(\eta_j) \le C.\end{aligned}$$

Therefore, we can use the result shown in Theorem 1 to solve the above problem. As shown in our theoretical study, the sequence $g(\beta^{(m)}) = \frac{1}{2n}\|Y - X\beta^{(m)}\|_2^2$ is decreasing and the sequence $\{\beta^{(m)}\}$ converges to a near optimal solution.

High-dimensional Cost-constrained Regression (HCR)

Step 1: Choose $\delta > 0$, $L > \ell = \lambda_{max}(\frac{1}{n}\mathbf{X}^T\mathbf{X})$, and initialize $\beta^{(1)}$ such that $\sum_{j=1}^p c_j \mathcal{I}(\beta_j^{(1)}) \leq C$.

Step 2: For $m \geq 1$, denote $\mu^{(m)} = \beta^{(m)} + \frac{1}{nL}\mathbf{X}^T(Y - \mathbf{X}\beta^{(m)})$, apply dynamic programming to find

$$\beta^{(m+1)} \in \arg\min_{\eta} \|\eta - \mu^{(m)}\|_2^2 \text{ subject to } \sum_{j=1}^p c_j \mathcal{I}(\eta_j) \leq C$$

$$= \mu^{(m)} \circ Z^{(m)} = (\mu_1^{(m)} z_1^{(m)}, \mu_2^{(m)} z_2^{(m)}, \ldots, \mu_p^{(m)} z_p^{(m)}), \text{ where}$$

$$Z^{(m)} \in \arg \max_{z_1, z_2, \ldots, z_p} \sum_{j=1}^p (\mu_j^{(m)})^2 z_j \text{ subject to } \sum_{j=1}^p c_j z_j \leq C.$$

Step 3: Repeat **Step 2**, until $g(\beta^{(m)}) - g(\beta^{(m+1)}) \leq \delta$.

Extension 1: p is much larger than n and β^0 is sparse

We aim to develop a linear model that will

- satisfy the budget constraint;
- have good prediction performance.

Since some feasible models may have k $(k > n)$ variables, we need to use regularization techniques to solve the overfitting problem.

We propose to estimate the regression coefficient vector by

$$\hat{\beta} \in \arg\min_{\beta} \frac{1}{2n} \|Y - \mathbf{X}\beta\|_2^2 + \sum_{j=1}^{p} \lambda(\alpha|\beta_j| + \frac{1-\alpha}{2}\beta_j^2)$$

$$\text{subject to } \sum_{j=1}^{p} c_j \mathcal{I}(\beta_j) \leq C.$$

Extension 1: p is much larger than n and β^0 is sparse

How to solve the optimization problem?

If $\hat{\beta}$ is an optimal solution to the following problem

$$\min_{\beta} \frac{1}{2}\|\beta - a\|_2^2 + \sum_{j=1}^{p} \lambda(\alpha|\beta_j| + \frac{1-\alpha}{2}\beta_j^2) \quad \text{subject to} \quad \sum_{j=1}^{p} c_j \mathcal{I}(\beta_j) \leq C,$$

then $\hat{\beta} = \frac{1}{1+\lambda(1-\alpha)} \cdot \mathrm{sign}(a - \alpha\lambda) \circ (|a| - \alpha\lambda)_+ \circ \hat{Z}$,
where $\hat{Z} = (\hat{z}_1, \hat{z}_2, \ldots, \hat{z}_p)$ is the solution to the following 0-1 knapsack problem

$$\max_{z_1, z_2, \ldots, z_p} \sum_{j=1}^{p} \frac{a_j^2 - 2\alpha\lambda|a_j| + \alpha^2\lambda^2}{2(1+\lambda(1-\alpha))} \cdot \frac{1 + \mathrm{sign}(|a_j| - \alpha\lambda)}{2} \cdot z_j$$

$$\text{subject to} \quad \sum_{j=1}^{p} c_j z_j \leq C, \quad \text{and } z_1, z_2, \ldots, z_p \in \{0, 1\}.$$

Extension 2: Convex differential loss functions with Lipschitz continuous gradient

Denote $f = \sum_{j=1}^{p} x_j\beta_j$ and let $\psi(y, f)$ be the loss function used to fit the model.

If the gradient of the convex differential loss function $\psi(y, f)$ satisfis the following Lipschitz condition

$$|\frac{\partial\psi}{\partial f}(y, f_1) - \frac{\partial\psi}{\partial f}(y, f_2)| \leq M_1|f_1 - f_2|,$$

for any y, f_1, f_2, and a positive constant M_1 (e.g., the squared hinge loss),

OR

$\frac{\partial\psi^2(y,f)}{\partial f^2}$ exists and $\frac{\partial\psi^2(y,f)}{\partial f^2} \leq M_2$ for any y and f, and a positive constant M_2 (e.g., the logistic regression loss),

if $L \geq 2M_1 \cdot \lambda_{max}(\frac{\mathbf{X}^T\mathbf{X}}{n})$ or $L \geq M_2 \cdot \lambda_{max}(\frac{\mathbf{X}^T\mathbf{X}}{n})$, we can also show that

$$g(\eta) \leq Q_L(\eta, \beta) = g(\beta) + \frac{L}{2}\|\eta - \beta\|_2^2 + \langle \nabla g(\beta), \eta - \beta \rangle,$$

for all β, η with equality holding at $\beta = \eta$.

Extension 3: Predictors are collected group-by-group

Suppose that we have G groups. For the g-th group, we need to spend $\tilde{c}_g$ dollars to simultaneously collect the values of all p_g variables in the g-th group $\mathcal{A}_g$.

We propose to estimate the regression coefficient vector by

$$\hat{\beta} \in \arg\min_{\beta} \frac{1}{2n}\|Y - \mathbf{X}\beta\|_2^2 + \sum_{j=1}^{p} \lambda(\alpha|\beta_j| + \frac{1-\alpha}{2}\beta_j^2)$$

$$\text{subject to } \sum_{g=1}^{G} \tilde{c}_g[1 - \prod_{j \in \mathcal{A}_g} (1 - \mathcal{I}(\beta_j))] \leq C.$$

Note that we assume that we always need to spend $\tilde{c}_g$ dollars if there is at least one variable in the g-th group $\mathcal{A}_g$ with a nonzero regression coefficient.

Extension 3: predictors are collected group-by-group

If $\hat{\beta}$ is an optimal solution to the following problem

$$\min_{\beta} \|\beta - a\|_2^2 \quad \text{subject to} \quad \sum_{g=1}^{G} \tilde{c}_g [1 - \prod_{j \in \mathcal{A}_g} (1 - \mathcal{I}(\beta_j))] \leq C,$$

then $\hat{\beta} = a \circ \hat{Z}$ where $\circ$ denotes the entrywise product of two vectors, $\hat{Z} = (\hat{z}_1 \mathbf{1}_{p_1}, \hat{z}_2 \mathbf{1}_{p_2}, \ldots, \hat{z}_g \mathbf{1}_{p_g})^T$, $\mathbf{1}_{p_g}$ is a row vector of p_g 1's, and $\hat{z}_1, \hat{z}_2, \ldots, \hat{z}_g$ is the solution to the following 0-1 knapsack problem

$$\max_{z_1, z_2, \ldots, z_g} \sum_{g=1}^{G} (\sum_{j \in \mathcal{A}_g} a_j^2) z_g$$

$$\text{subject to} \quad \sum_{g=1}^{G} \tilde{c}_g z_g \leq C, \ \text{and} \ z_1, z_2, \ldots, z_g \in \{0, 1\}.$$

Extension 3: predictors are collected group-by-group

If $\hat{\beta}$ is an optimal solution to the following problem

$$\min_{\beta} \frac{1}{2}\|\beta - a\|_2^2 + \sum_{j=1}^{p} \lambda(\alpha|\beta_j| + \frac{1-\alpha}{2}\beta_j^2)$$

$$\text{subject to } \sum_{g=1}^{G} \tilde{c}_g[1 - \prod_{j \in \mathcal{A}_g} (1 - \mathcal{I}(\beta_j))] \leq C,$$

then $\hat{\beta} = \frac{1}{1+\lambda(1-\alpha)} \cdot \text{sign}(a - \alpha\lambda) \circ (|a| - \alpha\lambda)_+ \circ \hat{Z}$,
where $\hat{Z} = (\hat{z}_1 \mathbf{1}_{p_1}, \hat{z}_2 \mathbf{1}_{p_2}, \ldots, \hat{z}_g \mathbf{1}_{p_g})^T$ and $\hat{z}_1, \hat{z}_2, \ldots, \hat{z}_g$ is the solution to the following 0-1 knapsack problem

$$\max_{z_1, z_2, \ldots, z_g} \sum_{g=1}^{G} (\sum_{j \in \mathcal{A}_g} \frac{a_j^2 - 2\alpha\lambda|a_j| + \alpha^2\lambda^2}{2(1+\lambda(1-\alpha))} \cdot \frac{1 + \text{sign}(|a_j| - \alpha\lambda)}{2}) z_g$$

$$\text{subject to } \sum_{g=1}^{G} \tilde{c}_g z_g \leq C, \text{ and } z_1, z_2, \ldots, z_g \in \{0, 1\}.$$

Theoretical Properties

(a) For any $L > \ell = \lambda_{max}(\frac{\mathbf{X}^T\mathbf{X}}{n})$, the sequence

$$g(\beta^{(m)}) = \frac{1}{2n}\|Y - \mathbf{X}\beta^{(m)}\|_2^2$$

is decreasing, converges and satisfies

$$g(\beta^{(m)}) - g(\beta^{(m+1)}) \geq \frac{L-\ell}{2}\|\beta^{(m+1)} - \beta^{(m)}\|_2^2.$$

Since the function $g(\beta)$ is nonnegative and convex, we know that the proposed algorithm will converge.

(b) If $L > \ell = \lambda_{max}(\frac{\mathbf{X}^T\mathbf{X}}{n})$, then

$$\beta^{(m+1)} - \beta^{(m)} \to 0 \text{ as } m \to \infty.$$

Theoretical Properties

Definition 1: Given an $L \geq \lambda_{max}(\frac{\mathbf{X}^T\mathbf{X}}{n})$, the vector $\eta \in R^p$ is said to be a first-order stationary point of our optimization problem

$$\min_{\beta} g(\beta) = \frac{1}{2n}\|Y - \mathbf{X}\beta\|_2^2 \text{ subject to } \sum_{j=1}^{p} c_j \mathcal{I}(\beta_j) \leq C$$

if

$$\sum_{j=1}^{p} c_j \mathcal{I}(\eta_j) \leq C$$

and

$$\eta = \arg\min_{\beta} \|\beta - (\eta - \frac{1}{L}\nabla g(\eta))\|_2^2 \text{ subject to } \sum_{j=1}^{p} c_j \mathcal{I}(\beta_j) \leq C$$

Theoretical Properties

(c) If $\hat{\beta}$ is a global minimizer of the optimization problem

$$\min_{\beta} g(\beta) = \frac{1}{2n}\|Y - \mathbf{X}\beta\|_2^2 \text{ subject to } \sum_{j=1}^{p} c_j \mathcal{I}(\beta_j) \le C,$$

then it is a first-order stationary point.

(d) If η is a first-order stationary point and

$$(\max_{i \in A} c_i) + \sum_{i \in A^c} c_i \le C,$$

where $A = \{i : \eta_i = 0\}$, then η is a global minimizer.

Theoretical Properties

(e) If $L > \lambda_{max}(\frac{\mathbf{X}^T\mathbf{X}}{n})$ and

$$\lim_{m\to\infty} \inf_{1\leq j\leq p} \min \{\max\{|\beta_j^{(m)}|, 0\}\} > 0,$$

then

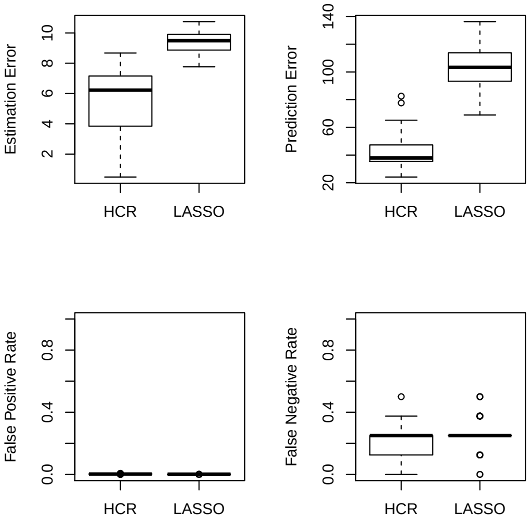
- The sparsity pattern sequence $Z^{(m)}$ converges after finitely many steps, that is, there exists an iteration index M^* such that $Z^{(m)} = Z^{(m+1)}$ for all $m \geq M^*$.
- The sequence $\beta^{(m)}$ is bounded and converges to a first-order stationary point.

The above theoretical properties about our HCR algorithm are similar to the theoretical results shown in Bertsimas et al. (2016) for the best subset selection.

Simulation Study: Example 1

- 200 training samples, 10000 testing samples;
- The dimension $p = 1000$ and $(x_1, x_2, \ldots, x_p)^T \sim N(0, \mathbf{I}_p)$;
- The model parameter β^0: the first 10 elements are randomly generated from $N(4, 1)$ and the other elements are 0;
- The costs of different variables are randomly selected from $\{1, 2, 3\}$. They are used for all the 100 experiments.
- The budget is $C = 12$ and the variance $\sigma^2 = 0.25$. Since the generated costs satisfy $\sum_{j=1}^{10} c_j = 18 > C$, the parameter of interest β^* is different from β^0.

Simulation Study: Example 1



Simulation Study: Example 2

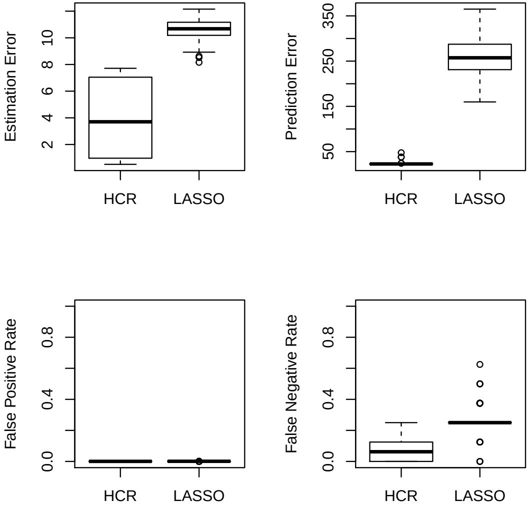
- The predictors $(x_1, x_2, \ldots, x_{10})^T \sim N(0, \mathbf{A})$ where $a_{jt} = 0.2$ if $j \neq t$ and 1 otherwise. The other $p - 10$ predictors are generated from $N(0, \mathbf{B})$ where $b_{jt} = 0.2$ if $j \neq t$ and 1 otherwise.

- The model parameter β^0: the first 10 elements are randomly generated from $N(4, 1)$ and the other elements are 0;

- The costs of different variables are randomly selected from $\{1, 2, 3\}$. They are used for all the 100 experiments.

- The budget is $C = 12$ and the variance $\sigma^2 = 0.25$. Since the generated costs satisfy $\sum_{j=1}^{10} c_j = 18 > C$, the parameter of interest β^* is different from β^0.

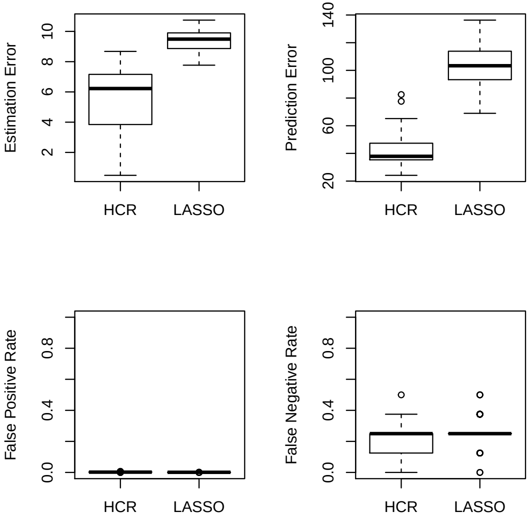
Simulation Study: Example 2



Simulation Study: Example 3

- The predictors $(x_{i1}, x_{i2}, \ldots, x_{ip})^T \sim N(0, \mathbf{\Sigma})$ where $\sigma_{jt} = 0.5^{|j-t|}$.
- The model parameter β^0: the first 10 elements are randomly generated from $N(4, 1)$ and the other elements are 0;
- The costs of different variables are randomly selected from $\{1, 2, 3, \ldots, 10\}$. They are used for all the 100 experiments.
- The budget is $C = 100$ and the variance $\sigma^2 = 0.25$. As C is equal to the cost of collecting all important variables, for this example, the true parameter of interest β^* is the same as β^0.

Simulation Study: Example 3



Application to a diabetes study

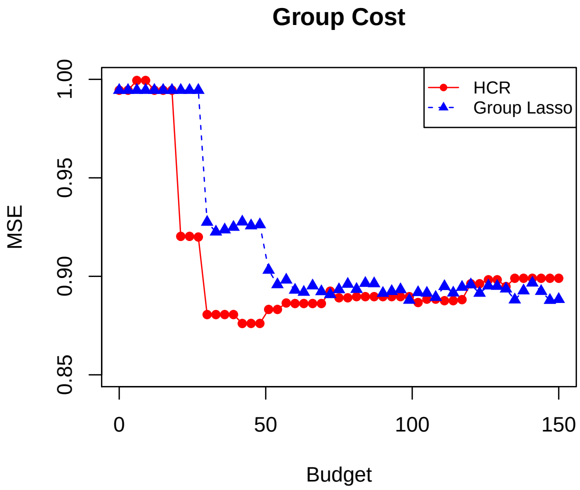
- The glycated hemoglobin (HbA1c) is now recommended as a standard of care (SOC) for testing and monitoring diabetes, specifically the type 2 diabetes.
- In general, the higher the HbA1c, the greater the risk of developing diabetes-related complications.
- We are interested in developing a cost-effective model to predict the change in HbA1c using some demographical information and predictors from several clinical tests.

Application to a diabetes study

- Our data are collected from 181 diabetes patients;
- The efficacy endpoint is the change in HbA1c from baseline to week 52;
- There are 20 predictors collected from 10 groups;
- We develop cost-constrained models for a sequence of budgets.

Costs	Predictors
$200	HDL
	LDL
	Total Cholesterol
	Triglycerides
$50	Creatinine
$20	Fasting BG
$30	HbA1c
$100	ALT
	AST
$100	GGT
	C.Peptide
$50	Fasting insulin test
$20	Age
	Weight
	BMI
	Waist
$5	Duraton of diabetes
$10	BP diastolic
	BP Systolic
	Pulse

Application to a diabetes study



Summary

- The cardinality budget constraint makes the high-dimensional cost-constrained regression problem NP-hard.
- We propose a new discrete extension of the first-order continuous optimization methods to deliver a near optimal solution.
- Our HCR algorithm generates a series of estimates of the regression coefficient vector by solving a sequence of 0-1 knapsack problems that can be efficiently addressed by the dynamic programming.
- The proposed HCR method can be extended to the other statistical learning problems and problems with more complicated constraints.

Thank you!