Optimal Large-Scale Internet Media Selection

Gareth James

Department of Data Sciences and Operations
Marshall School of Business
University of Southern California

October 21st, 2016

Joint with Paat Rusmevichientong, Lan Luo and Courtney Paulson
Consider the following constrained, penalized optimization problem:

$$\arg\min_{\beta} g(\beta) + \lambda \|\beta\|_1 \quad \text{subject to} \quad C\beta = b,$$

where $g$ is a convex function, $\beta \in \mathbb{R}^p$, $C \in \mathbb{R}^{m \times p}$ and $b \in \mathbb{R}^m$ are predefined matrices and vectors.

There turn out to be many important problems that can be formulated in this fashion so we are interested in algorithms for solving (1) and the statistical properties of the resulting coefficient estimates.
Maximizing Reach Or Click Through Rate

- The reach of an advertising campaign is defined as the probability a random customer views our ad at least once during the campaign, while click through rate (CTR) is the probability a random customer clicks the ad.
Maximizing Reach Or Click Through Rate

- The reach of an advertising campaign is defined as the probability a random customer views our ad at least once during the campaign, while click through rate (CTR) is the probability a random customer clicks the ad.

- We have $p$ websites and an advertising budget of $B$. Let $\beta_j$ be the allocation to the $j$th website and $g(\beta)$ represent either the estimated non-reach (or non-CTR) for a given budget allocation.
Maximizing Reach Or Click Through Rate

- The reach of an advertising campaign is defined as the probability a random customer views our ad at least once during the campaign, while click through rate (CTR) is the probability a random customer clicks the ad.

- We have $p$ websites and an advertising budget of $B$. Let $\beta_j$ be the allocation to the $j$th website and $g(\beta)$ represent either the estimated non-reach (or non-CTR) for a given budget allocation.

- Hence, we wish to minimize $g(\beta)$ such that $\sum_j \beta_j \leq B$ and $\beta_j \geq 0$. Or equivalently minimize $g(\beta) + \lambda \|\beta\|_1$.
The reach of an advertising campaign is defined as the probability a random customer views our ad at least once during the campaign, while click through rate (CTR) is the probability a random customer clicks the ad.

We have $p$ websites and an advertising budget of $\$B$. Let $\beta_j$ be the allocation to the $j$th website and $g(\beta)$ represent either the estimated non-reach (or non-CTR) for a given budget allocation.

Hence, we wish to minimize $g(\beta)$ such that $\sum_j \beta_j \leq B$ and $\beta_j \geq 0$. Or equivalently minimize $g(\beta) + \lambda \|\beta\|_1$.

However, in many campaigns we also wish to place restrictions on subsets of websites e.g. a cruise operator may wish to spend 30% of their budget on travel websites. This imposes a natural constraint of the form $C\beta = b$. 
A Quadratic Approximation

- We can approximate $g$ by

$$g(\beta) \approx \frac{1}{2} \| Y - X \beta \|^2_2 + K$$

where $X = H^{1/2}$, $Y = H^{-1/2}(J - H\tilde{\beta})$, $H$ is the Hessian and $J$ is the Jacobian (both evaluated at $\tilde{\beta}$).

- Hence, (1) can be approximated using the constrained lasso criterion. Minimize,

$$\arg \min_{\beta} \frac{1}{2} \| Y - X \beta \|^2_2 + \lambda \| \beta \|_1 \quad \text{subject to} \quad C\beta = b, \quad (2)$$
Suppose that we are given an index set, \( A \), of size \( m \).
Suppose that we are given an index set, \( \mathcal{A} \), of size \( m \).

Then we can partition

\[
\beta = (\beta_\mathcal{A}, \beta_{\overline{\mathcal{A}}}), \quad C = (C_\mathcal{A}, C_{\overline{\mathcal{A}}}).
\]

Hence,

\[
C_\mathcal{A}\beta_\mathcal{A} + C_{\overline{\mathcal{A}}}\beta_{\overline{\mathcal{A}}} = b
\]
Suppose that we are given an index set, $\mathcal{A}$, of size $m$. Then we can partition

$$\beta = (\beta_A, \beta_{\bar{A}}), \quad C = (C_A, C_{\bar{A}}).$$

Hence,

$$C_A \beta_A + C_{\bar{A}} \beta_{\bar{A}} = b$$

Or

$$\beta_A = C_A^{-1} (b - C_{\bar{A}} \beta_{\bar{A}}).$$
Hence, all we need to do is compute

$$\beta_{\bar{A}} = \arg \min_{\theta} \frac{1}{2} \| Y^* - X^* \theta \|^2_2 + \lambda \| \theta \|_1 + \lambda \| C_{\bar{A}}^{-1} (b - C_{\bar{A}} \theta) \|_1,$$

and set $\beta_{\bar{A}} = C_{\bar{A}}^{-1} (b - C_{\bar{A}} \beta_{\bar{A}}).$
Hence, all we need to do is compute

$$\beta_{\bar{A}} = \arg\min_{\theta} \frac{1}{2} \| Y^* - X^* \theta \|^2 + \lambda \| \theta \|_1 + \lambda \| C_{\bar{A}}^{-1} (b - C_{\bar{A}} \theta) \|_1,$$

and set $\beta_{\bar{A}} = C_{\bar{A}}^{-1} (b - C_{\bar{A}} \beta_{\bar{A}})$.  

The difficulty in computing (3) lies in the non-differentiability and non-separable nature of the second $\ell_1$ penalty.
Intuition

However, if we choose an $m$-vector, $s$, such that

$$s = \text{sign} (\beta_A),$$

then for $\theta$ close enough to $\beta_{\bar{A}}$

$$\| C_{\bar{A}}^{-1} (b - C_{\bar{A}} \theta) \|_1 = s^T C_{\bar{A}}^{-1} (b - C_{\bar{A}} \theta)$$

and we can replace the $\ell_1$ penalty by a differentiable term which no longer needs to be separable.
Intuition

- However, if we choose an \(m\)-vector, \(s\), such that
  \[
  s = \text{sign} (\beta_A),
  \]
  then for \(\theta\) close enough to \(\beta_A\)
  \[
  \|C_A^{-1} (b - C_A \theta)\|_1 = s^T C_A^{-1} (b - C_A \theta)
  \]
  and we can replace the \(\ell_1\) penalty by a differentiable term which no longer needs to be separable.

- Now our optimization becomes
  \[
  \beta_A = \arg \min_{\theta} \frac{1}{2} \|Y^* - X^* \theta\|_2^2 + \lambda s^T C_A^{-1} (b - C_A \theta) + \lambda \|\theta\|_1
  \]
  \[
  = \arg \min_{\theta} \frac{1}{2} \|\tilde{Y} - \tilde{X} \theta\|_2^2 + \lambda \|\theta\|_1.
  \]
Toy Example
Select Index Set $\mathcal{A}$ for $m = 2$
Check Coefficients in $A$ Maintained Same Sign

Coefficients have not crossed zero so solution is correct.
Not Every Index Set Will Work

Coefficients have not crossed zero so solution is correct.

\( \lambda_0 \)

This coefficient crossed zero so would have caused a problem to use.

\( \lambda_1 \)
Select New Index Set for Next Step

New index set.
Click Through Rate

![CTR (Full)](chart1.png)

![CTR (Travel Subset)](chart2.png)
Summary

- A large number of real world problems are special cases of this constrained and penalized framework.
- A simple algorithm, using standard lasso fitting methods, can be used to efficiently compute the solution to our optimization problem.
- Theoretical bounds on the coefficients can be extended from the lasso and suggest better performance.
- Simulation results show practical improvement, computational efficiency and relative insensitivity to the constraints.
- Provides a highly efficient and practical approach to select optimal allocations of advertising budget in situations involving thousands of websites.