Group LARS and Group Lasso

Xiang Ni

August 29, 2017

Overview

Initialization

- Regression coefficient: $\beta^{[0]} = 0$
- residual: $r^{[0]} = Y$.
- compute the current 'most correlated set'

$$A_1 = \mathrm{argmax}_j ||X_j^T r^{[0]}||^2 / p_j,$$

i.e. the factor that has the smallest angle with Y.

Inside the loop

For each step, the algorithm does three things:

- calculate step size α.
- decide which group factor to add.
- update regression coefficient $\beta^{[k]} = \beta^{[k-1]} + \alpha\gamma$ and residual $r^{[k]} = Y - X\beta^{[k]}$.

The program terminates when all the group factors have been added.

Direction vector

Before doing the first two things, we need to calculate the "direction vector" γ which is a $p = \sum p_j$ dimensional vector with $\gamma_{A_k^c} = 0$ and

$$\gamma_{A_k} = (X_{A_k}^T X_{A_k})^{-1} X_{A_k}^T r^{[k-1]},$$

i.e. the regression coefficient decided by factors in A_k.

How to calculate the step size

Question: what is step size?
Answer: For each $j \notin A_k$, step size α_j is how far the group LARS algorithm will progress in direction γ before X_j enters the most correlated set:

$$||X_j'(r^{[k-1]} - \alpha_j X\gamma)||^2/p_j = ||X_{j'}^T(r^{[k-1]} - \alpha_j X\gamma)||^2/p_{j'},$$

where j' is arbitrarily chosen from A_k.

calculate step size and choose which factor to add

Then

$$\alpha = \min_{j \notin A_k}(\alpha_j) = \alpha_{j^*}.$$

Update $A_{k+1} = A \cup \{j^*\}$.

How to calculate the step size

In the case of $k = 1$:

$$||X_j'(Y - \alpha_j X\gamma)||^2/p_j = ||X_{j'}^T(Y - \alpha_j X\gamma)||^2/p_{j'},$$

i.e.

$$||X_j' r^{[1]}||^2/p_j = ||X_{j'}^T r^{[1]}||^2/p_{j'},$$

In other works, the step size is found via proceeding in the direction of the projection Y on the space that is spanned by the factor until some other factor (i.e. j') has as small an angle with the current residual.

More thoughts

1. At this point the project of the current residual on the space that is spanned by the columns of X_j and $X_{j'}$ has equal angle with the two factors.
2. Next step: the group LARS proceeds in the direction of angular bisector.

Remarks of the algorithms

1. If two group of factors are almost equally correlated with the residual, then their coefficients should increase at approximately the same rate (coincide with intuition).
2. Good thing: solution path is piecewise linear.
3. $\alpha_j \in [0, 1]$ is well defined by properties of continuous functions.
4. The algorithm stops after J steps, where J is the number of groups.

Questions

1. how to understand the role of p_j?
2. mathematically, why group LARS is good?
3. how to intuitively understand the group LARS algo ?

Questions

Running against the diabetes dateset:

- Fabient's group_lasso.py produces
  $[3.08437933, -191.32678864, 515.6595972, 281.31633277, -50.72195946, -0., -224.39715802, 0., 478.89888885, 36.22682728]$.
- my group_LARS.py produces (by setting all p_j as 1):
  $[-16.01387468, -232.60633569, 518.07490076, 315.16696747, 0., 0., -346.33982461, -110.43048864, 499.03994811, 67.44355675]$.

Group lasso

Problem:

$$\frac{q}{2}||Y - \sum_{j=1}^{J} X_j\beta_j||^2 + \lambda \sum_{j=1}^{J} ||\beta_j||_K j,$$

where

$$||\eta||_K = (\eta^T K\eta)^{1/2}.$$

KKT Solution

Let $K_j = p_j I_j$. Then β is a solution of Group lasso iff

$$-X_j^T(Y - X\beta) + \frac{\lambda \beta_j p_j^{1/2}}{||\beta_j||} = 0,$$

for all β_j not equal to 0, and

$$|| - X_j^T(Y - X\beta)|| \leq \lambda\sqrt{p_j}, \quad \forall \beta_j = 0.$$

KKT Solution

If $X_j^T X_j = I_{p_j}$, then the solution to the KKT conditions are:

$$\beta_j = (1 - \frac{\lambda\sqrt{p_j}}{||S_j||})_+ S_j,$$

where $S_j = X_j^T(Y - X\beta_{-j})$, with $\beta_{-j} = (\beta_1', ...\beta_{j-1}', 0, ..., \beta_J')$

KKT Solution

If X is orthonormal:

$$\beta_j = (1 - \frac{\lambda\sqrt{p_j}}{||X_j^T Y||})_+ X_j^T Y.$$