CrossValidated

GLMNET or LARS for computing LASSO solutions?

Asked 10 years, 7 months ago Active today Viewed 9k times

14

8

I would like to get the coefficients for the LASSO problem

$$||Y - X\beta|| + \lambda||\beta||_1.$$

The problem is that glmnet and lars functions give different answers. For the glmnet function I ask for the coefficients of $\lambda/||Y||$ instead of just λ, but I still get different answers.

Is this expected? What is the relationship between the lars λ and glmnet λ? I understand that glmnet is faster for LASSO problems but I would like to know which method is more powerful?

deps_stats I am afraid that the size of my dataset is so large that LARS can not handle it, whereas on the other hand glmnet can handle my large dataset.

mpiktas I want to find the solution of (Y-Xb)^2+L\sum|b_j| but when I ask from the two algorithms(lars & glmnet) for their calculated coefficients for that particular L, I get different answers...and I wondering is that correct/ expected? or I am just using a wrong lambda for the two functions.

r machine-learning regression lasso regularization

Share Cite Edit Follow Flag

edited Aug 8 '16 at 2:43
Glen_b
252k 29 532 913

asked Feb 10 '11 at 16:23
user3139

1 please provide an example illustrating your problem. Also how do you define power of algorithm? – mpiktas Feb 10 '11 at 17:11

1 I've used glmnet and lars packages in a couple of projects. In my limited experience I've had A LOT of problems implementing glmnet. I think that glmnet needs some bug fixes regarding the type of variables used in the data frame. Besides, glmnet has confusing documentation. I ended up using lars, and I was very satisfied with the results. Nevermind the size of your problem, I think lars can handle it. – deps_stats Feb 10 '11 at 19:44

"The problem is that glmnet and lars functions give different answers." i have the same problem. Any answers ? – grant Mar 19 '12 at 12:01

Drastically different answers for coefficients? And just from reading the original post, you really shouldn't ask for a single lambda solution from `glmnet` and likely not from a LARS implementation either. They provide a whole range of solutions along the spectrum of bias vs variance. Which makes it hard to compare actual coefficients. But still, the same variables should probably become non-zero in a similar order. – Shea Parkes Mar 19 '12 at 12:31

3 Answers

Active | Oldest | Votes

15

In my experience, LARS is faster for small problems, very sparse problems, or very 'wide' problems (much much more features than samples). Indeed, its computational cost is limited by the number of features selected, if you don't compute the full regularization path. On the other hand, for big problems, glmnet (coordinate descent optimization) is faster. Amongst other things, coordinate descent has a good data access pattern (memory-friendly) and it can benefit from redundancy in the data on very large datasets, as it converges with partial fits. In particular, it does not suffer from heavily correlated datasets.

The conclusion that we (the core developers of the scikit-learn) have come too is that, if you do not have strong a priori knowledge of your data, you should rather use glmnet (or coordinate descent optimization, to talk about an algorithm rather than an implementation).

Interesting benchmarks may be compared in Julien Mairal's thesis:

https://lear.inrialpes.fr/people/mairal/resources/pdf/phd_thesis.pdf

Section 1.4, in particular 1.4.5 (page 22)

Julien comes to slightly different conclusions, although his analysis of the problem is similar. I suspect this is because he was very much interested in very wide problems.

Share Cite Edit Follow Flag

edited 19 mins ago

community wiki
3 revs, 3 users 90%
Gael Varoquaux

Most of your responses are made CW (here, but also on metaoptimize.com/qa)... Any reason why? – chl Apr 12 '11 at 20:07

1 Because I think that it is good if people can fix typos, incorrect wordings... I like the idea of continuous improvements of answers. Is this against best practice? – Gael Varoquaux Apr 12 '11 at 20:17

2 I fixed two or three typos in passing. Nothing to do with CW per se, unless you're thinking of the lower rep required to edit your post! Users can suggest any edits, though; but making your response CW won't allow you to earn rep from them. I guess you're not after rep, but as your responses were always well put, I was just wondering... Cheers. – chl Apr 12 '11 at 20:48

1

Lars and Glmnet give different solutions for the Lasso problem, as they use slightly different objective functions and different standartisations of the data. You can find details code for reproduction in the related question Why do Lars and Glmnet give different solutions for the Lasso problem?

Share Cite Edit Follow Flag

edited Apr 13 '17 at 12:44
Community Bot
1

answered Aug 9 '12 at 8:55
Andre
561 3 8

LASSO is non-unique in the case where multiple features have perfect collinearity. Here's a simple thought experiment to prove it.

1

Let's say you have three random vectors y, x_1, x_2. You're trying to predict y from x_1, x_2. Now assume y = $x1$ = $x2$. An optimal LASSO solution would be $\beta_1 = 1 - P$, $\beta_2 = 0$, where P is the effect of LASSO penalty. However, also optimal would be $\beta_1 = 0$, $\beta_2 - 1 - P$.

Share Cite Edit Follow Flag

edited Feb 10 '11 at 20:30

answered Feb 10 '11 at 19:19

dsimcha
7,105 6 32 28

3 @dsmcha, sorry to say this, but I don't think I like that example too much. The response is identical to two of the predictors? That's beyond pathological, in my view. – cardinal Feb 11 '11 at 0:31