Mathematics > Statistics Theory

The Lasso Problem and Uniqueness

Ryan J. Tibshirani

(Submitted on 1 Jun 2012 (v1), last revised 4 Nov 2012 (this version, v2))

The lasso is a popular tool for sparse linear regression, especially for problems in which the number of variables p exceeds the number of observations n. But when p>n, the lasso criterion is not strictly convex, and hence it may not have a unique minimum. An important question is: when is the lasso solution well-defined (unique)? We review results from the literature, which show that if the predictor variables are drawn from a continuous probability distribution, then there is a unique lasso solution with probability one, regardless of the sizes of n and p. We also show that this result extends easily to ℓ_1 penalized minimization problems over a wide range of loss functions.
A second important question is: how can we deal with the case of non-uniqueness in lasso solutions? In light of the aforementioned result, this case really only arises when some of the predictor variables are discrete, or when some post-processing has been performed on continuous predictor measurements. Though we certainly cannot claim to provide a complete answer to such a broad question, we do present progress towards understanding some aspects of non-uniqueness. First, we extend the LARS algorithm for computing the lasso solution path to cover the non-unique case, so that this path algorithm works for any predictor matrix. Next, we derive a simple method for computing the component-wise uncertainty in lasso solutions of any given problem instance, based on linear programming. Finally, we review results from the literature on some of the unifying properties of lasso solutions, and also point out particular forms of solutions that have distinctive properties.

Comments:	25 pages, 0 figures
Subjects:	**Statistics Theory (math.ST)**
Cite as:	**arXiv:1206.0313 [math.ST]**
	(or **arXiv:1206.0313v2 [math.ST]** for this version)

Submission history

From: Ryan Tibshirani [view email]
[v1] Fri, 1 Jun 2012 21:34:10 GMT (29kb)
[v2] Sun, 4 Nov 2012 20:16:58 GMT (29kb)