



# Model Selection for High-Dimensional Regression under the Generalized Irrepresentability Condition

Adel Javanmard, Andrea Montanari

(Submitted on 2 May 2013)

In the high-dimensional regression model a response variable is linearly related to  $p$  covariates, but the sample size  $n$  is smaller than  $p$ . We assume that only a small subset of covariates is 'active' (i.e., the corresponding coefficients are non-zero), and consider the model-selection problem of identifying the active covariates. A popular approach is to estimate the regression coefficients through the Lasso ( $\ell_1$ -regularized least squares). This is known to correctly identify the active set only if the irrelevant covariates are roughly orthogonal to the relevant ones, as quantified through the so called 'irrepresentability' condition. In this paper we study the 'Gauss-Lasso' selector, a simple two-stage method that first solves the Lasso, and then performs ordinary least squares restricted to the Lasso active set. We formulate 'generalized irrepresentability condition' (GIC), an assumption that is substantially weaker than irrepresentability. We prove that, under GIC, the Gauss-Lasso correctly recovers the active set.

Comments: 32 pages, 3 figures

Subjects: **Statistics Theory (math.ST)**; Information Theory (cs.IT); Learning (cs.LG); Methodology (stat.ME); Machine Learning (stat.ML)

Cite as: [arXiv:1305.0355](#) [math.ST]  
(or [arXiv:1305.0355v1](#) [math.ST] for this version)

## Submission history

From: Adel Javanmard [[view email](#)]

[v1] Thu, 2 May 2013 07:25:52 GMT (2565kb,D)

*Which authors of this paper are endorsers?*

## Download:

- [PDF](#)
- [Other formats](#)

Current browse context:

math.ST

[< prev](#) | [next >](#)

[new](#) | [recent](#) | [1305](#)

Change to browse by:

cs

[cs.IT](#)

[cs.LG](#)

math

stat

[stat.ME](#)

[stat.ML](#)

## References & Citations

- [NASA ADS](#)

## Bookmark (what is this?)

