Accepted Manuscript

Broken adaptive ridge regression and its asymptotic properties

Linlin Dai, Kani Chen, Zhihua Sun, Zhenqiu Liu, Gang Li

PII: S0047-259X(17)30506-7
DOI: https://doi.org/10.1016/j.jmva.2018.08.007
Reference: YJMVA 4402

To appear in: *Journal of Multivariate Analysis*

Received date : 22 August 2017

Please cite this article as: L. Dai, K. Chen, Z. Sun, Z. Liu, G. Li, Broken adaptive ridge regression and its asymptotic properties, *Journal of Multivariate Analysis* (2018), https://doi.org/10.1016/j.jmva.2018.08.007

Broken adaptive ridge regression and its asymptotic properties

Linlin Dai[a], Kani Chen[b], Zhihua Sun[c], Zhenqiu Liu[d], Gang Li[e,*]

[a]*Southwestern University of Finance and Economics, Chengdu, China*
[b]*Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong*
[c]*Institute of Mathematics, Ocean University of China, Qingdao, China*
[d]*Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA*
[e]*Department of Biostatistics, School of Public Health, University of California at Los Angeles, CA 90095-1772, USA*

Abstract

This paper studies the asymptotic properties of a sparse linear regression estimator, referred to as broken adaptive ridge (BAR) estimator, resulting from an L_0-based iteratively reweighted L_2 penalization algorithm using the ridge estimator as its initial value. We show that the BAR estimator is consistent for variable selection and has an oracle property for parameter estimation. Moreover, we show that the BAR estimator possesses a grouping effect: highly correlated covariates are naturally grouped together, which is a desirable property not known for other oracle variable selection methods. Lastly, we combine BAR with a sparsity-restricted least squares estimator and give conditions under which the resulting two-stage sparse regression method is selection and estimation consistent in addition to having the grouping property in high- or ultrahigh-dimensional settings. Numerical studies are conducted to investigate and illustrate the operating characteristics of the BAR method in comparison with other methods.

Keywords: Feature selection, Grouping effect, L_0-penalized regression, Oracle estimator, Sparsity recovery

1. Introduction

Simultaneous variable selection and parameter estimation is an essential task in statistics and its applications. A natural approach to variable selection is L_0-penalized regression, which directly penalizes the cardinality of a model through well-known information criteria such as Mallow's C_p [35], Akaike's information criterion (AIC) [1], the Bayesian information criterion (BIC) [9, 38], and risk inflation criteria (RIC) [17]. It has also been shown to have an optimality property for variable selection and parameter estimation [40]. However, the L_0-penalization problem is nonconvex and finding its global optima requires exhaustive combinatorial best subset search, which is NP-hard and computationally infeasible even for data in moderate dimension. Moreover, it can be unstable for variable selection [6]. A popular alternative is L_1-penalized regression, or Lasso [43], which is known to be consistent for variable selection [36, 48] although not consistent for parameter estimation [13, 51]. During the past two decades, much efforts have been devoted to improving Lasso using various variants of the L_1 penalty, which are not only consistent for variable selection, but also consistent for parameter estimation [13, 15, 24, 27, 47, 51]. Yet another approach is to approximate the L_0 penalty using a surrogate L_0 penalty function, which could lead to a more sparse model with favorable numerical properties [11, 25, 40, 41].

This paper concerns a different L_0-based approach to variable selection and parameter estimation which performs iteratively reweighted L_2 penalized regressions in order to approximate an L_0 penalized regression [19, 33]. Unlike a surrogate L_0 penalization method such as the truncated Lasso penalization (TLP) method [41], the L_0-based iteratively reweighted L_2 penalization method can be viewed as performing a sequence of surrogate L_0 penalizations, where each reweighted L_2 penalty serves as an adaptive surrogate L_0 penalty and the approximation of L_0 penalization improves with each iteration. It is also computationally appealing since each iteration only involves a reweighted L_2

*Corresponding author
Email address: vli@ucla.edu (Gang Li)