Computational Statistics and Data Analysis 104 (2016) 66–78

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Integrative weighted group lasso and generalized local quadratic approximation

Qing Pan [a,*], Yunpeng Zhao [b]

[a] The Department of Statistics, George Washington University, Washington, DC, United States
[b] The Department of Statistics, George Mason University, Fairfax, VA, United States

HIGHLIGHTS

- Weighted group lasso emphasizing extremum selects dynamic biomarker effects.
- Parameters in the loss, penalty and weight functions are estimated simultaneously.
- Local quadratic approximation is generalized to non-convex optimization problems.

ARTICLE INFO

Article history:
Received 13 October 2015
Received in revised form 19 April 2016
Accepted 13 June 2016
Available online 27 June 2016

Keywords:
Adaptive group lasso
Integrative group lasso
Generalized local quadratic approximation
GWAS
Optimization of non-convex function
Varying-coefficient regression

ABSTRACT

Longitudinal clinical outcomes are often collected in genomic studies, where selection methods accounting for dynamic effects of biomarkers are desirable. Biomarker effects can be modeled by nonparametric B-splines and selected by group lasso. A novel weight function is proposed based on the extremum of the biomarker effects over time for the penalty. In addition to the common practice treating weights as adaptive functions depending on some first-stage estimates, an integrative group lasso which treats the loss, penalty and weight functions as an integrative whole is proposed, where parameters in all three are jointly estimated in one step. Generalized local quadratic approximations are developed to optimize the integrative group lasso whose guidelines are applicable in a wide range of non-convex optimization problems. The integrative version has theoretical advantages as it requires weaker assumptions in achieving consistency and sparsistency. Both adaptive and integrative procedures show larger areas under the ROC curves as well as smaller biases and mean square prediction errors over unweighted group lasso in simulation studies. Finally, the proposed method is illustrated on the GWAS from the Epidemiology and Intervention of Diabetes Complication trial. To accommodate more candidate markers, 23 chromosomes are analyzed separately with common tuning parameters.

1. Introduction

In the presence of large number of candidate markers with inherent group structure, group lasso (Yuan and Lin, 2006) or COSSO (Lin and Zhang, 2006) are desirable for their group sparsity property. However, similar to lasso, regression coefficients estimates from group lasso are biased. By contrast, Wei et al. (2011) proved that under some regularity conditions, group lasso with adaptive weights selects the correct subset of variables with probability converging to one. Besides adaptive

* Corresponding author. Fax: +1 202 994 6517.
E-mail address: qpan@gwu.edu (Q. Pan).

http://dx.doi.org/10.1016/j.csda.2016.06.004
0167-9473/

weights where parameters in the weight function are estimated in the first stage and plugged into the penalized log-likelihood in the second stage, we propose integrative weights, which are functions of the unknown parameter values. In the adaptive version, the regression coefficient estimates depend on the accuracy of the estimates in the first stage. In the integrative group lasso, the parameters in the weight function are set to be unknown, and hence the loss function and the weighted penalty function are minimized jointly in one step as a unified whole. Such a weight function does not require any prior information on or first-stage estimation of the parameters. As a consequence, the integrative group lasso has theoretical advantages—it achieves sparsistency under milder conditions. However, neither traditional local quadratic approximation (Fan and Li, 2001) nor group LARS (Yuan and Lin, 2006) can solve the integrative group lasso directly due to the unknown parameters in the weights. We develop a generalized local quadratic approximation (GLQA) that gives a convex quadratic approximation of any penalty function and is guaranteed to converge in combination with back-tracking line search (Conway, 2004). The three guidelines of GLQA, which were not found in the literature by our knowledge, can be borrowed into other non-convex optimization problems.

The comparison of the adaptive and integrative weights also provides new insights into the adaptive lasso (Zou, 2006) and the smoothly clipped absolute deviation (SCAD) (Fan and Li, 2001) methods. Although appearing to be different, adaptive lasso and SCAD both modify the lasso penalty to achieve asymptotic consistency in both selection and estimation. The difference lays in the form of their penalty functions—adaptive weights depend on the first-stage estimators while SCAD penalty is a function of the unknown parameters. The same difference lays between the adaptive group lasso and our integrative group lasso. The adaptive and integrative group lasso procedures modify group lasso following the spirits of adaptive lasso and SCAD, respectively, in the following sense: Adaptive group lasso and adaptive lasso share the same essential properties; while both the integrative group lasso and SCAD construct penalties which are complicated functions of the unknown parameters, resulting in selection and estimation procedures with desirable properties such as unbiasedness and sparsity.

Our study is motivated by a genome wide association study (GWAS) with longitudinal clinical outcomes, where the effects of SNPs are presumably time-varying. Various nonparametric splines have been widely employed in literature to model time-varying effects of biomarkers. Take a few examples: Fan et al. (2012) used linear penalized splines to model temporal trends of SNP effects on longitudinal quantitative traits in analyzing the Framingham Heart Study Genetic Analysis Workshop data. Wang et al. (2007) also modeled the effects of transcription factors as B-splines. Chen and Zhang (2008) used multivariate adaptive splines to describe the relationship between the presence of regulatory motifs and gene expression. Chen and Wang (2011) estimated functional mixed effects models in which both the random and fixed effects are modeled as P-splines. Yan and Huang (2002) extended the adaptive group lasso procedure to Cox proportional hazards models and selected variables by maximizing penalized partial likelihood where the two penalties represented time-invariant effects and time-varying effects, respectively.

By using the adaptive and integrative lasso, we propose novel variable selection and coefficient estimation procedures for time-varying effects with an emphasis on effects with large extremum. The effects of the markers over time are modeled as cubic B-splines. And all the coefficients in constructing the same spline are viewed as a group, and selected or unselected jointly by group lasso with adaptive or integrative weights. Furthermore, we design novel weight functions based on the largest absolute values of the spline coefficients in the group. The weights based on extremum are motivated from, but not limited to, the following biological scenarios—when the true marker effects are small values around zero, they are most likely biological fluctuations without serious disease consequences. In such cases, the proposed method will assign large penalties, leading to lower chances of selection. On the contrary, we target at markers with large effects for at least a period of time, which pass the threshold to trigger disease onset or progression. In summary, the proposed method prefers markers with large temporary effects for at least a period of time over markers with consistently small fluctuations.

The rest of the article is organized as follows. The adaptive and integrative versions of the weighted group lasso are introduced in Section 2. Section 3 describes the computation algorithms for both versions. Asymptotic selection consistency and estimation consistency of the parameters are derived in Section 4, with emphasis on the difference between the conditions required by adaptive and integrative weights. Section 5 examines their performances and compares them to the unweighted group lasso through simulation studies. Finally, the motivating data are analyzed by both procedures in Section 6.

2. Weighted group lasso based on extremum

First we introduce the notation. Define $Y = [y_{ik}]$ as an $n \times T$ matrix, where y_{ik} is the outcome for the ith subject at the kth time point. Outcomes are measured at $t_1, \ldots, t_T$. Let $X = [x_{ip}]$ be an $n \times (P+1)$ matrix. The ith row $X_{i\cdot}$ denotes the covariate vector for the ith subject, where $x_{i0} = 1$ corresponds to the intercept and $x_{i1}, \ldots, x_{iP}$ represent candidate biomarkers. Note that the index p starts from 0. We assume $(y_{i1}, \ldots, y_{iT}, x_{i1}, \ldots, x_{iP})$ are independently and identically distributed for all i. The relationship between Y and X is modeled as,

$$y_{ik} = \sum_{p=0}^{P} x_{ip}\beta_p(t_k) + \epsilon_{ik}, \quad \epsilon_{ik} \text{ i.i.d. } \sim N(0, \sigma^2),$$