Statistical Methodology 32 (2016) 147–160

Contents lists available at ScienceDirect

Statistical Methodology

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

Going beyond oracle property: Selection consistency and uniqueness of local solution of the generalized linear model

Chi Tim Ng [a,*], Seungyoung Oh [b], Youngjo Lee [b]

[a] *Department of Statistics, Chonnam National University, Gwangju 500-757, South Korea*

[b] *Department of Statistics, Seoul National University, Seoul 151-747, South Korea*

ARTICLE INFO

Article history:
Received 22 September 2015
Received in revised form
23 May 2016
Accepted 26 May 2016
Available online 8 June 2016

MSC:
62F12
62J12

Keywords:
Generalized linear model
Penalized likelihood estimation
Oracle property
SCAD penalty
Selection consistency

ABSTRACT

Recently, the selection consistency of penalized least square estimators has received a great deal of attention. For the penalized likelihood estimation with certain non-convex penalties, search space can be constructed within which there exists a unique local minimizer that exhibits selection consistency in high-dimensional generalized linear models under certain conditions. In particular, we prove that the SCAD penalty of Fan and Li (2001) and a new modified version of the unbounded penalty of Lee and Oh (2014) can be employed to achieve such a property. These results hold even for the non-sparse cases where the number of relevant covariates increases with the sample size. Simulation studies are provided to compare the performance of SCAD penalty and the newly proposed penalty.

1. Introduction

The penalized likelihood approach based on the least absolute shrinkage and selection operator (LASSO) penalty of Tibshirani [18] has been widely used in the simultaneous variable selection and estimation. One advantage of the LASSO estimation is the uniqueness of the local solution due to the

* Corresponding author.
E-mail addresses: easterlyng@gmail.com (C.T. Ng), mnbv1689@naver.com (S. Oh), youngjo@snu.ac.kr (Y. Lee).

http://dx.doi.org/10.1016/j.stamet.2016.05.006

convexity. Such a unique LASSO solution exhibits selection consistency under the "irrepresentable condition" in [23]. However, this condition is too stringent to hold in practice.

As alternatives to the LASSO penalty, many non-convex penalties have been proposed. However, the uniqueness of the local solution is less trivial. To tackle this difficulty, various attempts have been made to define and study the selection consistency for the non-convex penalties. Fan and Li [2] introduce the oracle property that at least one of the local solutions is consistent with the true model and proposed the smoothly clipped absolute deviation (SCAD) penalty. Recently, Lee and Oh [14] propose a random-effect model that leads to a new unbounded (UB) penalty. The oracle property of the resulting estimator under the high-dimensional generalized linear models is established by Kwon et al. [9]. However, such an oracle property does not necessarily imply the selection consistency of the local solution that is actually obtained numerically.

Even though the local solutions are not necessarily unique, the local solution obtained from certain numerical procedure can be selection consistent. Fan, Xue, and Zou [6] show that the solution obtained by their local linear approximation algorithm allows selection consistency. Wang, Kim, and Li [19] show that the SCAD solution obtained by the concave–convex procedure (CCCP) of Yuille and Anand [20] is selection consistent under the linear regression models. Other kinds of approximations to the local solutions and global solution are also discussed in [22]. However, it remains unclear if such results hold for the solutions obtained from other numerical algorithms.

Loh and Wainwright [15] and Loh and Wainwright [16] consider the estimation consistency of the local solutions under a general class of penalized objective function encompassing the penalized sum of squares as a special case. Though error bounds are given, such results only guarantee that the coefficients of the irrelevant covariates go to zero, but not exactly zero.

The oracle property does imply the selection consistency if the local solution is unique. Kim and Kwon [8] and Zhang and Zhang [22] investigate the sufficient conditions for the uniqueness of local solutions. However, their results are obtained under the linear regression model. An objective of this paper is to generalize these uniqueness results to the penalized likelihood of generalized linear model (GLM). The global uniqueness of the local solution is difficult to prove for the GLM. However, we show in this paper that search space $\aleph$ can be chosen inside which, the local solution to the penalized likelihood estimation of the GLMs is unique and selection consistent. Here, selection consistency refers to exact zero estimates in the irrelevant covariates. Due to the uniqueness, the selection consistency does not rely on the particular algorithm. Such results are applicable to the SCAD penalized likelihood estimators of the GLMs under certain conditions on the design matrix. Furthermore, by modifying the UB penalty function of Lee and Oh [13] and Lee and Oh [14], the selection consistency and uniqueness of local solution can also be established. Our results hold even in the non-sparse cases where k, the number of relevant covariates increases with n, the sample size under the high-dimensional GLMs. The modified UB penalty allows a larger search space $\aleph$ than SCAD penalty to guarantee the selection consistency of the unique local solutions. Simulation studies further suggest that the newly proposed modified unbounded (MUB) penalty tends to require relatively smaller sample size n to achieve selection consistency, compared with the SCAD.

The remainder of this paper is organized as follows. In Section 2, the penalized likelihood estimator is defined. In Section 3, theorems on the selection consistency and uniqueness of penalized likelihood solutions with the SCAD Fan and Li [2] and a modified version of UB penalties are provided. In Sections 4 and 5, a simulation study is presented and an example using real data is given. Concluding remarks are given in Section 6. The proofs are given in Appendix A.

2. Penalty functions

Let $(\boldsymbol{x}_t, y_t)$, $t = 1, 2, \ldots, n$ be the observations, where $\boldsymbol{x}_t$ is the p-dimensional covariate vector and y_t is the scalar response. Consider GLMs with $E(y_t) = \mu_t$ and Var $(y_t) = \phi V(\mu_t)$, where ϕ is the dispersion parameter and $V(\cdot)$ is the variance function. The variance function characterizes the distribution of the GLM family with the density function

$$f(y_t; \mu_t) = \exp[\{\theta_t y_t - b(\theta_t)\}/\phi + c(y_t, \phi)],$$