Computational Statistics and Data Analysis 101 (2016) 289–299

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

l_1 regularized multiplicative iterative path algorithm for non-negative generalized linear models

B.N. Mandal [a,*], Jun Ma [b]

[a] ICAR-Indian Agricultural Statistics Research Institute, Pusa, New Delhi 110012, India

[b] Department of Statistics, Macquarie University, New South Wales 2109, Australia

ARTICLE INFO

Article history:
Received 14 August 2015
Received in revised form 19 March 2016
Accepted 24 March 2016
Available online 1 April 2016

Keywords:
Generalized linear models
Lasso
Elastic net
l_1-norm penalty
Regularization path
Non-negativity constraints

ABSTRACT

In regression modeling, often a restriction that regression coefficients are non-negative is faced. The problem of model selection in non-negative generalized linear models (NNGLM) is considered using lasso, where regression coefficients in the linear predictor are subject to non-negative constraints. Thus, non-negatively constrained regression coefficient estimation is sought by maximizing the penalized likelihood (such as the l_1-norm penalty). An efficient regularization path algorithm is proposed for generalized linear models with non-negative regression coefficients. The algorithm uses multiplicative updates which are fast and simultaneous. Asymptotic results are also developed for the constrained penalized likelihood estimates. Performance of the proposed algorithm is shown in terms of computational time, accuracy of solutions and accuracy of asymptotic standard deviations.

1. Introduction

For generalized linear models (GLMs), a random variable Y follows a distribution belonging to the exponential family and its mean value μ is related to the linear predictor $\eta = \mathbf{x}^\top \boldsymbol{\beta}$, where $\mathbf{x}$ is a vector for the data of a set of explanatory variables and $\boldsymbol{\beta}$ denotes a vector for the regression coefficients. Usually, one assumes that the mean and the linear predictor are linked through a link function g, i.e., $g(\mu) = \eta$. The regression coefficients are estimated by maximizing the log-likelihood:

$$\widehat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \log L(\boldsymbol{\beta}; \mathbf{X}, \mathbf{y}), \tag{1}$$

where $\log L$ denotes the log-likelihood derived from the data vector $\mathbf{y}$ and $\mathbf{X}$ is a matrix combining all the $\mathbf{x}$'s.

Lasso was proposed by Tibshirani (1996) for simultaneously performing variable selection and shrinkage for least-squares regressions when some important variables are to be selected out of p predictor variables. Lasso can be conceived as adopting a penalty using the l_1-norm of the regression coefficients, where the l_1-norm penalty can efficiently enforce sparsity. This concept was extended to GLMs by Park and Hastie (2007) by adding an l_1 penalty to the log-likelihood function. Without loss of generality, we assume that the predictor variables are standardized, i.e., $\sum_{i=1}^n x_{ij} = 0$ and $\sum_{i=1}^n x_{ij}^2 = 1$ for $j = 1, 2, \ldots, p$. Then, l_1 regularized estimates in GLM are given by

$$\widehat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \{\log L(\boldsymbol{\beta}; \mathbf{X}, \mathbf{y}) - \lambda \|\boldsymbol{\beta}\|_1\}, \tag{2}$$

* Corresponding author. Tel.: +91 11 25847121; fax: +91 11 25841564.
E-mail address: mandal.stat@gmail.com (B.N. Mandal).

http://dx.doi.org/10.1016/j.csda.2016.03.009
0167-9473/

where $\lambda > 0$ is the regularization parameter. Park and Hastie (2007) proposed an algorithm to compute the entire regularization path of (2) using a predictor corrector method. Friedman et al. (2010) developed a highly efficient coordinate descent method to solve (2).

Often, situations may arise that one needs the regression coefficients to be non-negative, such as non-negative least squares (NNLS) which has been studied intensively in the regression literature; see Lawson (1995) and Franc et al. (2005). Further motivating examples of non-negativity regression coefficients can be found in, for example, McDonald and Diamond (1990), Sha and Lin (2007) and Wu et al. (2013).

Recently, non-negative lasso has been considered by Wu et al. (2013) where the regression coefficients are constrained to be non-negative. They have used a multiplicative updating algorithm suggested by Sha and Lin (2007) for non-negative quadratic programming problems. However, their algorithm is not readily applicable to l_1-norm (or other) penalized GLMs with non-negative coefficients since the corresponding log-likelihood functions are not necessarily quadratic in general. If the penalty function contains only l_1-norm, such as in expression (2), then the coordinate decent algorithm of Friedman et al. (2010) can be modified to accommodate non-negative coefficients. This method, however, is difficult to be extended to more complex penalty functions, such as the fused lasso of Tibshirani et al. (2005). Other limitations of coordinate descent include, for example, that it can be neither parallelized nor implemented in a block-wise way—two important approaches to accelerate computation and convergence.

We propose in this paper a multiplicative iterative algorithm for l_1 penalized GLMs with non-negative coefficients. This multiplicative iterative algorithm has already been implemented in other applications, such as image processing (Chan and Ma, 2012), baseline hazard function estimation in Cox regression (Ma et al., 2014b) and medical image reconstruction (Ma, 2010). This algorithm is easy to implement as it involves only the first derivative and it has been shown to converge much faster when compared with, for example, some splitting algorithms (which can be conceived as special alternating direction minimization method (ADMM)) in image processing (Chan and Ma, 2012). A parallel or block-iterative version of this algorithm can be easily developed. For the purpose of simplifying discussions, we use only l_1 penalty when explaining this algorithm, but it can be easily extended to more complicated penalties. Particularly, the extension to the case of non-negative elastic net is also discussed in this paper. Another important contribution of this paper is that we develop asymptotic results for constrained penalized likelihood estimation. We will demonstrate accuracy of the asymptotic variance formula through a simulation study.

This paper is outlined as follows. In Section 2 the details of the multiplicative iterative algorithm are presented for estimation of parameters in l_1 penalized GLMs subject to non-negativity constraints on the regression coefficients. Section 3 discusses the elastic net extension of this method but still within the GLM framework. Asymptotic results for constrained lasso are presented in Section 4. Section 5 demonstrates the efficiency of this algorithm by reporting its computation time and other simulation results. A brief discussion on selection of the regularization parameter is presented in Section 6. We apply the proposed algorithm to a real data in Section 7. Section 8 concludes the article.

2. Multiplicative iterative algorithm for l_1 penalized non-negative GLMs

In this section, we describe the details of the multiplicative iterative algorithm to obtain the regularization path, namely solving (2) for various values of λ under the non-negativity constraints. The constrained optimization of our interest is

$$\widehat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}\geq\mathbf{0}}\{\phi(\boldsymbol{\beta}) = l(\boldsymbol{\beta}; \mathbf{X}, \mathbf{y}) - \lambda\|\boldsymbol{\beta}\|_1\}, \tag{3}$$

where $l = \log L$ and the constraint $\boldsymbol{\beta} \geq \mathbf{0}$ is interpreted elementwise. Since $\boldsymbol{\beta} \geq \mathbf{0}$, the objective function in (3) becomes

$$\phi(\boldsymbol{\beta}) = l(\boldsymbol{\beta}; \mathbf{X}, \mathbf{y}) - \lambda\boldsymbol{\beta}^\top \mathbf{1} \tag{4}$$

where $\mathbf{1}$ denotes a vector of ones with the same size as $\boldsymbol{\beta}$ and the superscript $\top$ denotes matrix transpose. The Karush–Kuhn–Tucker necessary conditions for the constrained optimization problem (3) are

$$\frac{\partial\phi}{\partial\beta_j} = 0 \quad \text{if } \beta_j > 0 \quad \text{and} \quad \frac{\partial\phi}{\partial\beta_j} < 0 \quad \text{if } \beta_j = 0. \tag{5}$$

Therefore, we wish to solve the following simultaneous equations

$$\beta_j\left(\frac{\partial l}{\partial\beta_j} - \lambda\right) = 0, \tag{6}$$

where $j = 1, 2, \ldots, p$ and subject to all $\beta_j \geq 0$. In (6) and for the GLM model considered by this paper, $\partial l/\partial\beta_j = \mathbf{x}_j^\top\mathbf{W}(\mathbf{y}-\boldsymbol{\mu})$.

McCullagh and Nelder (1989), where $\mathbf{x}_j = (x_{j1}, \ldots, x_{jn})^\top$ is the vector of n observations on the jth predictor variable, $\mathbf{W}$ is an $n \times n$ diagonal matrix with the ith diagonal element $1/(V_i\partial\eta_i/\partial\mu_i)$ with $V_i = \mathrm{var}(Y_i)$ and $(\mathbf{y} - \boldsymbol{\mu})$ is a vector with n elements $(y_i - \mu_i)$.

We propose to solve (6) by utilizing the Multiplicative Iterative (MI) algorithm developed by Ma (2006); see also Ma (2010) and Chan and Ma (2012) for different applications of this algorithm. Writing $\partial l/\partial\beta_j = (\partial l/\partial\beta_j)^+ + (\partial l/\partial\beta_j)^-$, where