Computational Statistics and Data Analysis 71 (2014) 347–358

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Fast Bayesian model assessment for nonparametric additive regression☆

S. McKay Curtis, Sayantan Banerjee*, Subhashis Ghosal

North Carolina State University, United States

ARTICLE INFO

Article history:
Received 30 May 2012
Received in revised form 15 April 2013
Accepted 15 May 2013
Available online 22 May 2013

Keywords:
Group LASSO
Laplace approximation
Model uncertainty
Penalized regression
Variable selection

ABSTRACT

Variable selection techniques for the classical linear regression model have been widely investigated. Variable selection in fully nonparametric and additive regression models has been studied more recently. A Bayesian approach for nonparametric additive regression models is considered, where the functions in the additive model are expanded in a *B*-spline basis and a multivariate Laplace prior is put on the coefficients. Posterior probabilities of models defined by selection of predictors in the working model are computed, using a Laplace approximation method. The prior times the likelihood is expanded around the posterior mode, which can be identified with the group LASSO, for which a fast computing algorithm exists. Thus Markov chain Monte-Carlo or any other time consuming sampling based methods are completely avoided, leading to quick assessment of various posterior model probabilities. This technique is applied to the high-dimensional situation where the number of parameters exceeds the number of observations.

1. Introduction

The literature abounds in variable selection methods for the linear model; see, for example, Miller (2002) and George (2000). One particular method that has generated a substantial amount of research is the Least Absolute Shrinkage and Selection Operator or LASSO (Tibshirani, 1996). This method involves minimizing penalized sums of squares where the penalty is the sum of the absolute values of the coefficients. For certain values of a tuning parameter, the minimizer of this penalized sum of squares can set one or more coefficients exactly to zero, and thus remove those variables from the model. A fast computing algorithm for the LASSO is given by a modification of the Least Angle Regression (LARS) algorithm (Efron et al., 2004). Many other variable selection approaches are variations on this penalized regression theme and typically differ from the LASSO by varying the form of the penalty; see, for example, Breiman (1995), Fan and Li (2001), Zou and Hastie (2005), Zou (2006), Bondell and Reich (2008), Hwang et al. (2009) and so on.

In many practical applications, the linear model setting is too restrictive and nonparametric regression models are preferred. In the recent years, several authors have proposed variable selection techniques for fully nonparametric regression. Friedman (1991) uses a forward stepwise regression procedure to construct a regression function from "reflected pairs" of basis functions. Linkletter et al. (2006) define the covariance function of a Gaussian process to be a function of individual predictors. Variables are selected by inclusion or exclusion from the covariance function. Lafferty and Wasserman (2008) use derivatives of the nonparametric function estimates with respect to smoothing parameters to find sparse solutions to the nonparametric variable selection problem.

☆ Research is partially supported by NSF grant number DMS - 1106570.

* Correspondence to: Department of Statistics, NCSU, 2311 Stinson Drive, SAS Hall, Raleigh, NC 27695-8203, United States. Tel.: +1 919 699 8773.
E-mail address: sbanerj5@ncsu.edu (S. Banerjee).

0167-9473/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.csda.2013.05.012

Although, fully nonparametric regression models are attractive in that they make relatively few assumptions about the regression function, they also lack the interpretability of the classical linear model. Additive models (Buja et al., 1989; Hastie and Tibshirani, 1990; Stone, 1985) provide a nice compromise between the restrictive linear model and the fully flexible nonparametric model. The additive model assumes that each predictor's contribution to the mean of the response can be modeled by an unspecified smooth function, thereby retaining some of the benefits of fully nonparametric regression. Additive models retain some of the benefits of interpretability found in classical linear models because each predictor has its own functional effect on the response. In addition, the simplifying assumptions of additive functional effects allow additive models to avoid the curse of dimensionality. Additive models can also be extended to smoothing-spline ANOVA (SS-ANOVA) models that allow for higher order interactions among the predictors (Barry, 1986; Gu, 2002; Wahba, 1990).

A handful of variable selection techniques exist for additive models. Chen (1993) develops a bootstrap procedure for model selection in SS-ANOVA models. Shively et al. (1999) develop a Bayesian model where the functional effect of each predictor is given a prior with a linear component and a nonlinear Wiener process component. Shi and Tsai (1999) give a modified version of Akaike's Information Criterion (AIC) (Akaike, 1974) suitable for selection of regression models with linear and additive components. Gustafson (2000) presents a Bayesian variable selection technique for regression models that allow predictors to have linear or functional effects and two-way interactions. Wood et al. (2002) develop a Bayesian method, based on the Bayesian Information Criterion (BIC) (Schwarz, 1978), for selecting between a linear regression model, a model with additive functional effects, or a fully nonparametric regression model. Lin and Zhang (2006) present the Component Selection and Smoothing Operator (COSSO) which is a generalization of the LASSO based on fitting a penalized SS-ANOVA model where the penalty is the sum of norms of the projection of each functional component into a partition of the model space. Belitz and Lang (2008) propose algorithms for variable selection and choosing the degree of smoothness in the regression models with structured additive predictors. Marra and Wood (2011) use shrinkage methods along with an extension of the non-negative garotte estimator for generalized additive models. Reich et al. (2009) develop a Bayesian variable selection technique for SS-ANOVA models with Gaussian process priors.

Yuan and Lin (2006) present a variable selection technique, called the group LASSO, for predictors that form natural groupings (e.g., sets of dummy variables for factors). Avalos et al. (2003) also develop a similar procedure for the special case of additive models using a B-spline basis. The group LASSO is a penalized least-squares method that uses a special form of penalty to eliminate redundant variables from the model simultaneously in pre-specified groups of variables. More specifically, let $\boldsymbol{Y}$ be an $n \times 1$ vector of responses, $\boldsymbol{X}_j$ is an $n \times m_j$ matrix of variables associated with the jth predictor (which may be stochastic or nonstochastic) and $\boldsymbol{\beta}_j$ is an $m_j \times 1$ vector of coefficients. Then group LASSO minimizes

$$\underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\| \boldsymbol{Y} - \sum_{j=1}^{g} \boldsymbol{X}_j \boldsymbol{\beta}_j \right\|^2 + \lambda \sum_{j=1}^{g} \|\boldsymbol{\beta}_j\|, \tag{1}$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1{}^T, \ldots, \boldsymbol{\beta}_g{}^T)^T$ and g is the number of groups. Yuan and Lin (2006) show that for some values of the tuning parameter λ, the solution to (1) includes $\boldsymbol{\beta}_j = \mathbf{0}$ for some subset of $j = 1, \ldots, g$. Other penalized approaches for variable selection in nonparametric additive regression models are also available (Meier et al., 2009; Ravikumar et al., 2009). Huang et al. (2010) consider variable selection in nonparametric additive regression models using a generalization of the adaptive LASSO (Zou, 2006) to the group LASSO (Yuan and Lin, 2006), called the adaptive group LASSO, and give conditions for consistent selection of the components in the underlying model.

One drawback of most variable selection methods is that they do not provide a measure of model uncertainty. Variable selection methods typically give one model as the best, without giving some measurement of uncertainty for this estimated model. The exceptions to this are methods that follow the Bayesian paradigm. They typically provide a measure of model uncertainty by calculating the number of times a particular model is visited in the posterior draws from a Markov chain Monte Carlo (MCMC) simulation (George and McCulloch, 1993). However, MCMC methods are computationally expensive when a large number of variables are involved and it can be hard to assess convergence when MCMC methods must traverse a space of differing dimensions. In fact, when the model dimension is very high, most MCMC based methods break down.

In this paper, we present a method for calculating approximate posterior model probabilities without having to draw MCMC samples. We use a multivariate Laplace prior on the coefficients of the functions in the model. In a linear model with normal errors, it is well known that when using independent univariate Laplace priors, the posterior mode coincides with the LASSO. Similarly the group LASSO can be viewed as the posterior mode with respect to some appropriate multivariate Laplace prior. The prior dependence in the components induces the grouping structure in the group LASSO. In additive models, we expand functions in a suitable basis such as the spline basis, and put a multivariate Laplace prior on the coefficients of the model. The coefficients of functions of the same predictors are taken to be a priori dependent, but coefficients of functions referring to different predictors are taken to be a priori independent. This introduces a natural grouping of variables formed by basis expansion of function of original predictor variables, for which the group LASSO is the posterior mode. We use the Laplace method of approximation of integrals by expanding the integrand around its maxima, thus avoiding costly MCMC simulations. Our method may be viewed as a generalization of the method of Yuan and Lin (2005), who develop a similar method for the classical linear regression model, by using the Laplace approximation around the standard LASSO. However, the main focus of Yuan and Lin (2005) was to obtain an empirical Bayes estimate of the tuning parameter of LASSO using the Bayesian approach. In contrast, our interest is truly in obtaining posterior probabilities of various models.