Journal of Multivariate Analysis 143 (2016) 71–93

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Bias-corrected inference for multivariate nonparametric regression: Model selection and oracle property

Francesco Giordano, Maria Lucia Parrella [*]

Department of Economics and Statistics, University of Salerno, Via Giovanni Paolo II, 132 - 84084 Fisciano (SA), Italy

ARTICLE INFO

Article history:
Received 30 July 2014
Available online 11 September 2015

AMS 2000 subject classification:
62G05
62G08
62G15

Keywords:
Multivariate nonparametric regression
Variable selection
Multivariate bandwidth selection
Multivariate confidence bands

ABSTRACT

The local polynomial estimator is particularly affected by the curse of dimensionality, which reduces the potential of this tool for large-dimensional applications. We propose an estimation procedure based on the local linear estimator and a sparseness condition that focuses on nonlinearities in the model. Our procedure, called BID (bias inflation–deflation), is automatic and easily applicable to models with many covariates without requiring any additivity assumption. It is an extension of the RODEO method, and introduces important new contributions: consistent estimation of the multivariate optimal bandwidth (the *tuning parameter* of the estimator); consistent estimation of the multivariate bias-corrected regression function and confidence bands; and automatic identification and separation of nonlinear and linear effects. Some theoretical properties of the method are discussed. In particular, we show the nonparametric oracle property. For linear models, BID automatically reaches the optimal rate $O_p(n^{-1/2})$, equivalent to the parametric case. A simulation study shows the performance of the procedure for finite samples.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

Let $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ be a set of $\mathbb{R}^{d+1}$-valued random vectors, where Y_i are the dependent variables and $\mathbf{X}_i$ are the $\mathbb{R}^d$-valued covariates of the model

$$Y_i = m(\mathbf{X}_i) + \varepsilon_i. \tag{1}$$

The function $m(\mathbf{X}_i) = E(Y_i|\mathbf{X}_i) : \mathbb{R}^d \to \mathbb{R}$ is the multivariate conditional mean function. The errors ε_i are assumed to be *i.i.d.* and independent of $\mathbf{X}_i$. We use the notation $\mathbf{X}_i = (X_i(1), \ldots, X_i(d))$ for the observed covariates and $\mathbf{x} = (x_1, \ldots, x_d)$ to denote the target point at which we want to estimate m. In addition, $f_X(\mathbf{x})$ is the density function of the covariate vector. It is assumed to be positive in $\mathbf{x}$ and continuous in a neighborhood of $\mathbf{x}$. Moreover, $f_\varepsilon(\cdot)$ is the density function of the errors, assumed to be $N(0, \sigma_\varepsilon^2)$.

Our goal is to estimate the function $m(\mathbf{x}) = E(Y|\mathbf{X} = \mathbf{x})$ at a point $\mathbf{x} \in \text{supp}(f_X)$ when the parametric form of the function m is completely unknown and we do not impose any additivity assumption. We assume that the number of covariates d is high, but that only some covariates are relevant. Analysis of this framework raises the problem of the dimensionality curse, which usually arises for nonparametric estimators, and consequently the problem of variable selection, which is a requirement for dimension reduction.

[*] Corresponding author.
E-mail address: mparrella@unisa.it (M.L. Parrella).

http://dx.doi.org/10.1016/j.jmva.2015.08.016

0047-259X/© 2015 Elsevier Inc. All rights reserved.

This nonparametric framework has been widely studied, as reviewed elsewhere [3]. For variable selection, approaches include penalty-based methods for semiparametric models [9,4] and neural network [8] and empirical [18] methods. Other strategies involve contextual variable selection and consistent estimation of the multivariate regression function, such as the COSSO [11], ACOSSO [17], and LAND [20] algorithms. All of these methods are appealing, but typical drawbacks include: difficulty in analyzing theoretically the properties of the estimators; the computational burden; the difficulty to implement the procedures, which generally depend crucially on some regularization parameters, quite difficult to set; the necessity of considering stringent assumptions on the functional space (e.g., imposing an additive model).

Here we propose a nonparametric multivariate regression method, called BID, based on the local linear estimator and its properties. It mediates among the following priorities: the need of being automatic, the need of scaling to high dimension and the need of adapting to large classes of functions. As far as we know, the first procedure that has met the above needs is the RODEO of [7]. The main finding of RODEO has been to solve the curse of dimensionality for multivariate kernel estimators without requiring the additivity assumption for the regression function. This is a property that we want to preserve in our procedure. Therefore, our work was inspired by RODEO [7] and is an extension of this approach. In particular, we borrow the idea of using an iterative procedure to "adjust" the multivariate estimation one dimension at a time. Moreover, we use the same test to identify nonlinearities. However, our procedure also introduces some interesting new contributions, described in the following.

First, the BID procedure includes a novel method for estimating the multivariate optimal bandwidth (i.e., the smoothing parameter of the estimator), which is completely automatic and easily applicable to models with many covariates. It is based on the assumption that each covariate can have a different bandwidth value. Note that bandwidths have a central role in the proposed procedure, since the values are used for variable selection and model selection.

Second, our procedure includes a consistent bias-corrected estimator for the multivariate regression function and the multivariate confidence bands. Bias is a significant problem in nonparametric kernel regression. Solving this problem is important for making precise inferences based on the estimated model. As far as we know, our procedure is the first one which is able to estimate the bias of the multivariate nonparametric estimator (including the sign of the bias, required to correct the estimation of the multivariate function).

Finally, our method automatically identifies and separates linearities and nonlinearities, as will be explained in Section 3. Moreover, BID has the nonparametric oracle property as defined elsewhere [17]: (i) it selects the correct subset of predictors with probability tending to one, and (ii) estimates the non-zero parameters as efficiently as if the set of relevant covariates were known in advance. This also fixes a drawback of RODEO: in fact, when RODEO is implemented with the local linear estimator, it is blind to the subset of linear predictors (those for which the partial derivative is constant with respect to the same predictor). Therefore, condition (i) of the nonparametric oracle property is not met in such a case. Actually, there are some solutions originally proposed by [7] to overcome this problem. In particular, they suggest to use the LASSO residual (to remove linearities first) or change the estimator with local constant smoothing (i.e., local polynomial of degree zero). But these solutions appear to be less attractive than ours, for several reasons that will be mentioned in Section 3.

We show that the rate of convergence of our final estimator is not sensitive to the number of relevant linear covariates in the model, even when the model is not additive. As a consequence, the effective dimension of the model can increase without incurring the curse of dimensionality, as long as the number of nonlinear covariates (those whose partial derivative is not constant with respect to the same covariate) remains fixed.

The remainder of the paper is organized as follows. Section 2 introduces the notation. The BID algorithm is presented in Section 3. In Section 4 we propose a method for estimating the multivariate optimal bandwidth, while Section 5 presents estimators of the functionals for deriving the bias-corrected multivariate confidence bands. Section 6 contains the main theoretical results. They are based on the fundamental assumption of uniform design for the covariates, as in [7]. In Section 7 we generalize the procedure to nonuniform frameworks. A simulation study concludes the paper. Assumptions and proofs are collected in the Appendix.

2. Basics and notation

The BID smoothing procedure is based on the local linear estimator (LLE), a nonparametric tool whose properties have been studied in depth [16]. It corresponds to a locally weighted least squares fit of a linear function, given by

$$\arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left\{ Y_i - \beta_0(\mathbf{x}) - \boldsymbol{\beta}_1^T(\mathbf{x})(\mathbf{X}_i - \mathbf{x}) \right\}^2 K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}), \tag{2}$$

where the function $K_{\mathbf{H}}(\mathbf{u}) = |\mathbf{H}|^{-1} K(\mathbf{H}^{-1}\mathbf{u})$ gives the local weights and $K(\mathbf{u})$ is the kernel function, a d-variate probability density function. The $d \times d$ matrix $\mathbf{H}$ represents the multivariate smoothing parameter, called the bandwidth matrix. It controls the variance of the kernel function and regulates the amount of local averaging on each dimension, and thus the local smoothness of the estimated regression function. We denote by $\boldsymbol{\beta}(\mathbf{x}) = (\beta_0(\mathbf{x}), \boldsymbol{\beta}_1^T(\mathbf{x}))^T$ the vector of coefficients to estimate at point $\mathbf{x}$. Using matrix notation, the solution of the minimization problem in (2) can be written in closed form as

$$\hat{\boldsymbol{\beta}}(\mathbf{x}; \mathbf{H}) = (\boldsymbol{\Gamma}^T \mathbf{W} \boldsymbol{\Gamma})^{-1} \boldsymbol{\Gamma}^T \mathbf{W} \boldsymbol{\Upsilon}, \tag{3}$$

Download English Version:

https://daneshyari.com/en/article/1145289

Download Persian Version:

https://daneshyari.com/article/1145289

Daneshyari.com