Local Adaptive Grouped Regularization and its Oracle Properties

Wesley Brooks, Jun Zhu, Zudi Lu

July 10, 2014

1 Introduction

Whereas the coefficients in traditional linear regression are scalar constants, the coefficients in a varying coefficients regression (VCR) model are functions - often *smooth* functions - of some effect-modifying variable (**??**).

Current practice for VCR models relies on global model selection to decide which variables should be included in the model, meaning that predictors are identified as relevant or irrelevant over the entire domain $\mathcal{D}$. **?** describe a method for globally selecting the relevant predictors in a VCR model where the coefficient functions are estimated with P-splines. **?** show a method for doing global variable selection in a VCR model where the coefficient functions are estimated by basis expansion.

Local adaptive grouped regularization (LAGR) is developed here as a method to select only the locally relevant predictors at any specific location $\boldsymbol{s}$ in the domain $\mathcal{D}$ of a VCR model. The method of LAGR applies to VCR models where the coefficients are estimated using locally linear kernel smoothing.

Using kernel smoothing for nonparametric regression is described in detail in **?**. The extension to estimating VCR models is made by **?** for a VCR a univariate effect-modifying variable, and by **?** for two-dimensional effect-modifying variable and autocorrelation among the obverved response. These methods minimize the boundary effect (**?**) by estimating the coefficients as local polynomials of odd degree (usually locally linear).

For linear regression models, the adaptive lasso (AL) (**?**) produces consistent estimates of the coefficients and has been shown to have appealing properties for automating variable selection, which under suitable conditions include the "oracle" property of asymptotically including exactly the correct set of covariates and estimating their coefficients as well as if the correct covariates were known in advance. For data where the obvserved variables fall into mutually exclusive groups that are known in advance, the adaptive group lasso (**??**) has similar oracle properties to the adaptive lasso while doing selection at the level of groups rather than individual variables. The proposed LAGR method uses the adaptive group lasso for local variable selection and coefficient estimation in a locally linear regression model. We show that LAGR posesses the oracle properties of asymptotically selecting exactly the correct local covariates and estimating their local coefficients as accurately as would be possible if the identity of the nonzero coefficients for the local model were known in advance.

The remainder of this document is organized as follows. The kernel-based VCR model is described in Section 2; the proposed LAGR technique and its oracle properties are presented in Section **??**; in Section **??**, the performance of the proposed LAGR technique is evaluated in a simulation study, and in Section **??** the proposed method is applied to the Boston house price dataset. Proofs of the theorems appear in Appendix **??**.

2 Varying coefficients regression

2.1 Model

Consider n data points, observed at sampling locations $\boldsymbol{s}_i = (s_{i,1},\ s_{i,2})^T$ for $i = 1, \ldots, n$, which are distributed in a spatial domain $\mathcal{D} \subset \mathbb{R}^2$ according to a density $f(\boldsymbol{s})$ with $\boldsymbol{s} \in \mathcal{D}$. For $i = 1, \ldots, n$, let $y(\boldsymbol{s}_i)$ and $\boldsymbol{x}(\boldsymbol{s}_i)$ denote, respectively, the univariate response and the $(p+1)$-variate vector of covariates measured at location $\boldsymbol{s}_i$. At each location $\boldsymbol{s}_i$, assume that the outcome is related to the covariates by a linear regression where the coefficients $\boldsymbol{\beta}(\boldsymbol{s}_i)$

are functions in two dimensions and $\varepsilon(\boldsymbol{s}_i)$ is random error at location $\boldsymbol{s}_i$. That is,

$$y(\boldsymbol{s}_i) = \boldsymbol{x}(\boldsymbol{s}_i)'\boldsymbol{\beta}(\boldsymbol{s}_i) + \varepsilon(\boldsymbol{s}_i). \tag{1}$$

Further assume that the error term $\varepsilon(\boldsymbol{s}_i)$ is normally distributed with zero mean and variance σ^2, and that $\varepsilon(\boldsymbol{s}_i)$, $i = 1, \ldots, n$ are independent. That is,

$$\boldsymbol{\varepsilon} \stackrel{iid}{\sim} \mathcal{N}\left(0, \sigma^2\right). \tag{2}$$

In the context of nonparametric regression, the boundary-effect bias can be reduced by local polynomial modeling, usually in the form of a locally linear model (**?**). Here, to prepare for the estimation of locally linear coefficients, we augment the local design matrix with covariate-by-location interactions in two dimensions (**?**). The augmented local design matrix at location $\boldsymbol{s}_i$ is

$$\boldsymbol{Z}(\boldsymbol{s}_i) = (\boldsymbol{X} \;\; \boldsymbol{L}_i\boldsymbol{X} \;\; \boldsymbol{M}_i\boldsymbol{X}) \tag{3}$$

where $\boldsymbol{X}$ is the unaugmented matrix of covariates, $\boldsymbol{L}_i = \text{diag}\{s_{i',1} - s_{i,1}\}$ and $\boldsymbol{M}_i = \text{diag}\{s_{i',2} - s_{i,2}\}$ for $i' = 1, \ldots, n$.

Now we have that $Y(\boldsymbol{s}_i) = \{\boldsymbol{Z}(\boldsymbol{s}_i)\}_i^T \boldsymbol{\zeta}(\boldsymbol{s}_i) + \varepsilon(\boldsymbol{s}_i)$, where $\{\boldsymbol{Z}(\boldsymbol{s}_i)\}_i^T$ is the ith row of the matrix $\boldsymbol{Z}(\boldsymbol{s}_i)$ as a row vector, and $\boldsymbol{\zeta}(\boldsymbol{s}_i)$ is the vector of