Investigation of Sparse Hierarchical Regularization for Basis Expansion Methods

Exploration and Expansion Regression via `HierBasis`

David Fleischer Annik Gougeon

Last Update: 02 May, 2018

Contents

1 Introduction

The method of nonparametric regression regularization described in Haris, Shojaie, and Simon (2016) provides a flexible framework and implementation of a sparse hierarchical penalty via the `R` package `HierBasis`. This proposal offered by outlines a convex penaltization and estimation technique that is suggested to be well-suited to high-dimensional problems. In particular, we wish to verify and expand upon the `HierBasis` framework in the context of sparse additive modelling, focusing on the problem of prediction of a continuous response and variable selection.

1.1 Problem Description

We restrict the focus of this project to focus on the problem of regression of a continuous response $y = [y_n, ..., y_n] \in \mathbb{R}^n$ on a high-dimensional design matrix $\mathbb{X} = [\boldsymbol{x}_1, ..., \boldsymbol{x}_n]^T = [X_1, ..., X_p] \in \mathbb{R}^{n \times p}$, such that

$$
\begin{aligned}
\boldsymbol{x}_i &= [x_{i1}, ..., x_{ip}] \quad \text{(observation } i\text{)} \\
X_j &= [x_{1j}, ..., x_{nj}]^T \quad \text{(predictor } j\text{)}.
\end{aligned}
$$

We consider the problem of estimating additive components $\{f_j\}_{j=1}^p$ of the additive model

$$y_i = \sum_{j=1}^p f_j\left(x_{ij}\right) + \varepsilon_i,$$

for a sparse set of active features embedded within the design matrix $\mathbb{X}$. The proposal offered by Haris, Shojaie, and Simon (2016) considers the class of basis expansion estimators (Cencov (1962)) defined by a finite set of basis functions $\{\psi_k(z)\}_{k=1}^K$, with some notion of increasing complexity (in k) and for a truncation level K to be adaptively selected. Let $\Psi_K^{(j)} \in \mathbb{R}^{n\times K}$ be the basis expansion corresponding to the j^{th} predictor X_j, with $(i,k)^{th}$ entry associated with observation x_{ij} and basis function ψ_k,

$$\Psi_{K,(i,k)}^{(j)} = \psi_k(x_{ij}), \quad 1 \le k \le K,\ 1 \le i \le n.$$

Then, through the basis expansion functions, the design matrix $\mathbb{X} \in \mathbb{R}^{n\times p}$ maps to a set of p $(n \times K)$ matrices

$$\mathbb{X} \overset{\psi}{\mapsto} \left\{\Psi_K^{(j)} \in \mathbb{R}^{n\times K}\right\}_{j=1}^p.$$

Of present interest is the set of polynomial basis functions $\{\psi_k(z)\}_{k=1}^K = \left\{z^k\right\}_{k=1}^K$ so that

$$X_j = \begin{bmatrix} x_{1j} \\ \vdots \\ x_{nj} \end{bmatrix} \mapsto \Psi_K^{(j)} = \begin{bmatrix} \psi_1(x_{1j}) & \psi_2(x_{1j}) & \cdots & \psi_K(x_{1j}) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(x_{nj}) & \psi_2(x_{nj}) & \cdots & \psi_K(x_{nj}) \end{bmatrix} = \begin{bmatrix} x_{1j} & x_{1j}^2 & \cdots & x_{1j}^K \\ \vdots & \vdots & \ddots & \vdots \\ x_{nj} & x_{nj}^2 & \cdots & x_{nj}^K \end{bmatrix}.$$

We estimate the additive components f_j by the <u>s</u>parse <u>a</u>dditive `HierBasis` estimator $\widehat{f}_j$ given by

$$\widehat{f}_j(x_{ij}) = \sum_{k=1}^K \widehat{\beta}_{j,k}^{\texttt{SA-hier}} \psi_k(x_{ij}), \quad j = 1, ..., p,$$

such that the $j = 1, ..., p$ coefficient vectors $\widehat{\boldsymbol{\beta}}_j^{\texttt{SA-hier}} = \left[\widehat{\beta}_{j,1}^{\texttt{SA-hier}}, ..., \widehat{\beta}_{j,K}^{\texttt{SA-hier}}\right] \in \mathbb{R}^K$ are simultaneously estimated by the minimization problem

$$\left[\widehat{\boldsymbol{\beta}}_1^{\texttt{SA-hier}}, ..., \widehat{\boldsymbol{\beta}}_p^{\texttt{SA-hier}}\right] = \underset{\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_p}{\arg\min} \left\{ \frac{1}{2n} \left\| Y - \sum_{j=1}^p \Psi_K^{(j)} \boldsymbol{\beta}_j \right\|_2^2 + \lambda \sum_{j=1}^p \Omega_j\left(\boldsymbol{\beta}_j\right) + \frac{\lambda^2}{\sqrt{n}} \sum_{j=1}^p \left\| \Psi_K^{(j)} \boldsymbol{\beta}_j \right\|_2 \right\}, \tag{1}$$

where

$$\Omega_j(\boldsymbol{\beta}_j) = \frac{1}{\sqrt{n}} \sum_{k=1}^K w_k \left\| \Psi_{k:K}^{(j)} \boldsymbol{\beta}_{j,k:K} \right\|_2,$$

for $w_k = k^m = (k-1)^m$ penalization weights for the k^{th}-order basis estimator, $\Psi_{k:K}^{(j)}$ denotes the submatrix of columns $k, k+1, ..., K$ of $\Psi_K^{(j)}$, and $\boldsymbol{\beta}_{j,k:K}$ denotes the corresponding subvector of $\boldsymbol{\beta}_j$.

The penalty described in (1) is defined by two terms. The first term containing Ω_j's is designed to provide a data-driven method of selecting the basis complexity/truncating the degree of the basis functions to some adaptively selected level $K_0 \le K$. This term is derived from the hierarchical group lasso penalty (Zhao, Rocha, and Yu (2009)) and leads to hierarchical sparsity of the fitted parameters. That is, $\widehat{\beta}_{j,k} = 0 \implies \widehat{\beta}_{j,k'} = 0$ for all $k' \ge k$.

The second term in the sparse additive `HierBasis` penalty $\frac{\lambda^2}{\sqrt{n}} \sum_{j=1}^p \left\| \Psi_K^{(j)} \boldsymbol{\beta}_j \right\|_2$ imposes sparsity across the predictors $X_1, ..., X_p$ and induces additional sparsity across the solution space $\left\{\widehat{\boldsymbol{\beta}}_j^{\texttt{SA-hier}}\right\}_{j=1}^p$.

1.2 Proposal

2 Methods

2.1 Proximal Methods for the HierBasis Penalty

2.2 The hierbasis2 Package

We have create a companion package to HierBasis, named hierbasis2, in order to implement the above tests and features of the above proposal. This new package retains all of the user-facing functionality of the original HierBasis package, but now permits the user to manipulate some additional parameters, as well as introducing some new functions. That is, the new library has been designed with this project in mind, allowing us to explore the properties of the original HierBasis package in a modular, readable, and concise format.

2.2.1 Installation

As is the case for HierBasis, installation of hierbasis2 can be done via devtools::install_github

```r
#install.packages(devtools)
library(devtools)
install_github("dfleis/hierbasis2")
library(hierbasis2)
```

3 Results

4 Discussion

References

Cencov, NN. 1962. "Estimation of an Unknown Density Function from Observations." In *Dokl. Akad. Nauk, Sssr*, 147:45–48.

Haris, Asad, Ali Shojaie, and Noah Simon. 2016. "Nonparametric Regression with Adaptive Truncation via a Convex Hierarchical Penalty." *arXiv Preprint arXiv:1611.09972*.

Zhao, Peng, Guilherme Rocha, and Bin Yu. 2009. "The Composite Absolute Penalties Family for Grouped and Hierarchical Variable Selection." *The Annals of Statistics*. JSTOR, 3468–97.