Journal of Statistical Software

MMMMMM YYYY, Volume VV, Issue II. *doi: 10.18637/jss.v000.i00*

The R Package BHAM: Fast and Scalable Bayeisan Hierarchical Additive Model for High-dimensional Data

Boyi Guo
University of Alabama at Birmingham

Nengjun Yi
University of Alabama at Birmingham

Abstract

\pkg{BHAM} is a freely avaible R pakcage that implments Bayesian hierarchical additive models for high-dimensional clinical and genomic data. The package includes functions that generlized additive model, and Cox additive model with the spike-and-slab LASSO prior. These functions implements scalable and stable algorithms to estimate parameters. \pkg{BHAM} also provides utility functions to construct additive models in high dimensional settings, select optimal models, summarize bi-level variable selection results, and visualize nonlinear effects. The package can facilitate flexible modeling of large-scale molecular data, i.e. detecting succeptable variables and inforing disease diagnostic and prognostic. In this article, we describe the models, algorithms and related features implemented in \pkg{BHAM}. The package is freely avaiable via the public GitHub repository https://github.com/boyiguo1/BHAM.

Keywords: additive model, spike-and-slab LASSO, scalable.

1. Introduction

Growing interest in fitting flexible and interpretable models, particular in high-dimensional data analysis, for the purpose of variable selection and predictive modelling.

Characteristic of these software includes: scalability, fast computation, and model flexibility and implementation easiness.

1.1. Literature Review

We enlist current available packages that have similar functionality, i.e. modeling to the best of our knowledge. To note, we don't list packages that are unable of handling high-

dimensional data, for example the well known R package `mgcv`, and high-dimensional packages that requires extra steps to construct the design matrix of functional form of predictors (Such implementation can be found with grouped sparse models, for example `SGL`.)

? Summarized the software development of additive models in high-dimensional data analysis before 2013.

Generalized Additive Model

- `COSSO`
- `spikeSlabGAM`
- `sparseGAM`

Additive Cox Proportional Hazard Model

- `COSSO`
- `tfCox`

The **BHAM** package provides a scalable solution for fitting high-dimensional generalized additive model and additive Cox model using spike-and-slab LASSO priors or other regularized priors, including continuous spike-and-slab priors, Student' T priors and double exponential priors. It fits linear, logistic, poisson and Cox regression models. The specification of the additive functions follows a popular syntax implemented in `mgcv`. Ancillary functions are provided, including cross-validation, model summary, and visualization.

In this article, we focus on the packages that can directly construct additive models for high-dimensional data analysis, instead of requiring additional step of constructing design matrix of functional form of the variables before fitting a sparse model.

There are other methods to model survival outcome and provides proporitonal hazards interpretation, for example **?** provides a link-based survival additive model for mixed censoring in package `GJRM`.

2. Bayesian Hierarchical Additive Model

2.1. Gerenalized Additive Model

2.2. Cox Proportional Hazard Model

3. R Functions

3.1. Model fitting

High-dimension Smoothing Formula

```
R> x <- 1:10
R> x
```

```
 [1]  1  2  3  4  5  6  7  8  9 10
```

Model Fitting

Covariate Adjustment

3.2. Model Summary

Functional Selection

Curve Plotting

Model Performance

4. Metabolomics Data Analysis with **BHAM**

4.1. Contious Outcome

4.2. Binary Outcome

4.3. Survival Outcome

5. Conclusion

This template demonstrates some of the basic LaTeX that you need to know to create a JSS article.

5.1. Code formatting

In general, don't use Markdown, but use the more precise LaTeX commands instead:

- Java

- **plyr**

One exception is inline code, which can be written inside a pair of backticks (i.e., using the Markdown syntax).

If you want to use LaTeX commands in headers, you need to provide a `short-title` attribute. You can also provide a custom identifier if necessary. See the header of Section 6 for example.

6. R code

Can be inserted in regular R markdown blocks.

6.1. Features specific to rticles

- Adding short titles to section headers is a feature specific to **rticles** (implemented via a Pandoc Lua filter). This feature is currently not supported by Pandoc and we will update this template if it is officially supported in the future.
- Using the `\AND` syntax in the `author` field to add authors on a new line. This is a specific to the `rticles::jss_article` format.

Affiliation:

Boyi Guo
University of Alabama at Birmingham
1665 University Blvd
Birmingham, AL 35294-0002 USA
E-mail: boyiguo1@uab.edu
URL: http://boyiguo1.github.io

Nengjun Yi
University of Alabama at Birmingham
1665 University Blvd
Birmingham, AL 35294-0002 USA
E-mail: nyi@uab.edu

Journal of Statistical Software http://www.jstatsoft.org/
published by the Foundation for Open Access Statistics http://www.foastat.org/
MMMMMM YYYY, Volume VV, Issue II *Submitted:* yyyy-mm-dd
doi:10.18637/jss.v000.i00 *Accepted:* yyyy-mm-dd