



COLLEGE OF PUBLIC HEALTH
DEPARTMENT OF BIostatISTICS
145 N. RIVERSIDE DRIVE, N300 CPHB
IOWA CITY, IOWA 52242

October 8, 2020

Dear editors,

We are submitting our manuscript “The biglasso Package: A Memory- and Computation-Efficient Solver for Lasso Model Fitting with Big Data in R” for your consideration as an original paper in *The R Journal*.

Our paper falls into two of the content categories listed on *The R Journal*’s website: “comparisons and benchmarking” and “add-on packages”. Current R packages for fitting lasso and elastic net models do not scale up well to very large data sets and crash when used with data larger than RAM. This motivated us to develop the **biglasso** package. The paper we are submitting describes this package and carries out a fair amount of benchmarking studies comparing it to existing R packages for accomplishing this task, such as **glmnet**.

Although we do provide a section with two examples of analyzing data with this package, we feel that the extensive benchmarking and comparison to other packages makes this paper suitable for publication in *The R Journal*. In particular, in addition to timing benchmarks, which are pretty common, we also provide memory usage benchmarks, which are often lacking in discussions of R packages, but are critical for software that intends to be capable of analyzing big data. Given the popularity of the lasso, we feel that this paper would be of interest to the readers of *The R Journal*.

The source code for this package is available on CRAN as well as GitHub:

- <https://cran.r-project.org/package=biglasso>
- <https://github.com/YaohuiZeng/biglasso>

Furthermore, the full source code to replicate all results from the manuscript is available in the following repository:

- https://github.com/YaohuiZeng/biglasso_reproduce

This work has not been previously published and is not under review elsewhere. This manuscript has been seen and approved by all authors. Thank you very much for your consideration.

Sincerely,

Patrick Breheny
Associate Professor
Department of Biostatistics