AMS 597: Statistical Computing

Pei-Fen Kuan (c)

Applied Math and Stats, Stony Brook University

Project Description

- Your project involves creating an R package that performs the following tasks.
- You will work in pairs (groups of 2)
- You can also opt out and work on the project individually (Deadline to opt out is Thursday March 10 at 5PM EST by emailing the instructor)
- The instructor will make initial group assignment randomly this Friday and post the pairing on Blackboard. You are allowed to swap groups
- Deadline for group swapping is March 24, 2022 at 10AM EST.
- Email the instructor your new group, cc'ing the members of your old and new group.

Project Description

- Your R package will take as input a response variable `y` and matrix of candidate predictors/independent variables `X`, where each column is a predictor.
- Your package will work for both binary `y` and continuous `y` (for continuous case, it can be assumed to be normally distributed).
- The number of predictors p can be large (i.e., you should also consider the case where $p > n$, n is the sample size).

Project Description

- Your package will allow user specifiy the type of model to fit:
  - linear or logistic regression
  - ridge regression (for binary and continuous `y`)
  - lasso regression (for binary and continuous `y`)
  - random lasso (for binary and continuous `y`)
- For random lasso, you will read the following paper: "Random Lasso", Annals of Applied Statistics (2011) `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3445423/`
- You will implement the algorithm described in Section 2 (Algorithm ("Generate" and "Select")).
- For Step 2b, you only need to implement for lasso (you do not need to implement adaptive lasso).
- The pdf version of this paper is available at Blackboard (see ScientificPaper_ProjectSpring2022.pdf)

Project Description

- Your function will fit a prediction model to the input data
- You can either treat the entire data as training data, or you can also make it more user friendly by allowing user to divide the data into training/test, and evaluate the model performance on the test data.
- Your package can import `glmnet` and use the functions in this package.
- You will then wrap these up as an R package called `extendedglmnet`

Project Description

- The R package has to be complete and contains a vignette describing how to use the R package
- The R package is due May 05, 2022 at 5:00 PM
- Submit your package as original source package (i.e., .tar.gz file) on Blackboard>Assignments>Project. Name your package `extendedglmnetGroupX_version.tar.gz`, where `GroupX` is your group number, e.g., `Group1`
- Version is generated automatically after you build your package successfully
- All students will submit the R package to Blackboard (i.e., although both members will submit the same R package, I still require both to submit to their respective Blackboard workspace).

Project Description

- Some of the grading criteria include:
  - Can the R package be installed successfully?
  - Is the R package implementing the required method correctly?
  - Has it considered all possible scenarios?
  - Is the R package user friendly (vignette, help files, warning messages, sample data, sample code)?
  - What is the computational speed?

Project Description

- Some useful links:
- https://tinyheero.github.io/jekyll/update/2015/07/26/making-your-first-R-package.html
- https://hilaryparker.com/2014/04/29/writing-an-r-package-from-scratch/
- https://combine-australia.github.io/r-pkg-dev/
- http://kbroman.org/pkg_primer/
- http://kbroman.org/Tools4RR/assets/lectures/08_rpack_withnotes.pdf
- https://cran.r-project.org/doc/contrib/Leisch-CreatingPackages.pdf
- https://ourcodingclub.github.io/tutorials/writing-r-package/

Project Description

- Some useful links for incorporating existing R package into your R package
- `https://kbroman.org/pkg_primer/pages/depends.html`
- `https://r-pkgs.org/description.html`
- Or google keywords `import R package''`,depends R package"
- Some useful links (for Windows):
- `https://www.biostat.wisc.edu/~kbroman/Rintro/Rwinpack.html`