Biostat 625 Homework #3

In this homework assignment, you need to develop an R package implementing an existing R function, or a group of R functions that are closely related to each other. Examples are linear regression model, logistic regression model, generalized linear models, Cox proportional hazards model, linear mixed model, support vector machine, etc. You can pick a topic of your interest, depending on your own level of statistical and computational knowledge. For your R package, you need to implement minimally the followings:

1. A complete R package, with R source code, documents (help pages), examples, etc.
2. The R package is hosted on a public GitHub repository, with a `README.md` file introducing the package.
3. Vignette(s) written in R markdown demonstrate the usage of the functions in the package.
4. In the vignette(s), include comparison(s) against the original R functions on simulated or real datasets to demonstrate both the correctness (e.g., via `all.equal()`) and the efficiency (e.g., via `bench::mark()`) of the implemented functions.

The following features are optional but highly recommended, and will be considered as bonuses during grading:

1. Include C++ code via `Rcpp`.
2. Include unit testing cases via `testthat`.
3. Include continuous integration.
4. Include code coverage test.

Many good examples of R packages can be found on GitHub, for instance, `https://github.com/tidyverse/dplyr`.

The following grading criteria will be used:

Required (for each item, 1: fail, 2: poor, 3: satisfactory, 4: good, 5: excellent)

1. The `README.md` is clearly written and sufficiently introduces the R package.
2. The help pages are clearly written and sufficiently introduce the function(s) in the R package.
3. The examples are clearly written and sufficiently demonstrate the function(s) in the R package, and can run smoothly without error or warning.
4. The R code is clearly written and well organized. The coding style is neat and consistent. The comments are clearly written and helpful.
5. The vignette(s) are clearly written and sufficiently demonstrate the usage of the function(s) in the R package.

6. The comparison(s) against the original R function(s) on simulated or real datasets clearly and sufficiently demonstrate both correctness and efficiency of the function(s) implemented in the R package.

7. Either sample dataset(s) included in the R package or a wide range of simulated data demonstrate the applications of the function(s) implemented in the R packages (e.g., via examples in help pages or in the vignette(s))

8. The amount of R code is non-trivial.

Optional (for each item, 1 bonus point if applicable)

1. C++ code is included via `Rcpp` and improves the efficiency of the function(s) implemented in the package (e.g., via comparison(s) against the original R functions(s)).

2. Unit testing cases are included in the R package via `testthat` and tests the function(s) implemented in the R package.

3. Continuous integration is included and integrated on the GitHub project site. All the tests are passed.

4. Code coverage test is included and integrated on the GitHub project site. The coverage is 100%.

Grading

Required (40 pt) + optional (up to 4 pt) = total (up to 40 pt).