Least Angle Regression and Shiny

Course Project of STAT 5361 Statistical Computing

*Biju Wang**

10/18/2018

Abstract

This project report will interactively visualize the leaset angle regression algorithm for fitting linear regression models to simulated data.

Least Angle Regression (LARS) is an algorithm for fitting linear regression models to high dimensional data. Suppose we expect a response variable to be determined by a linear combination of a subset of potential covariates. Then the LARS algorithm provides a means of producing an estimate of which variables to include, as well as their coefficients.

This algorithm is similar to forward stepwise regression. While forward stepwise regression builds a model sequentially, adding one variable at a time. At each step, it identifies the best variable to include in the active set and then updates the least square fit to include all the active variables. Least angle regression only enters "as much" of a predictor as it deserves. At the first step, it identifies the variable most correlated with the response. Rather than fit this variable completely, LAR moves the coefficient of this variable continuously toward its least squares value. As soon as another variable "catches up" in terms of correlation with the residual, the process is paused. The second variable then joins the active set, and their coefficients are moved together in a way that keeps their correlations tied and decreasing. This process is continued until all the variables are in the model and ends at the full least-squares fit. The whole procedure will terminate after $\min\{N-1, p\}$ steps where N is the number of observations and p is the size of covariates.

Shiny provides an incredible way to build interactive web applications. Shiny has automatic "reactive" binding between inputs and outputs and extensive pre-built widgets. It can also be embedded in R Markdown.

In this project, I will write LARS algorithm for simulated data and compare the outcomes from my codes with the ones from the lars package. Shiny apps will be used to provide an interactive interface for the users to see how each step the coefficients change by sliding a button or inputting step number. More complex interfaces may also be added as I further learn how to build Shiny apps.

References

1. Efron, B., Hastie, T., Johnstone, I., Tibshirani, R. (2004). Least Angle Regression. Annals of Statistics, 32(2).
2. Hastie, T., Tibshirani, R., Friedman, J. (2008). The Elements of Statistical Learning. New York, NY, USA: Springer New York Inc.
3. Shiny from R Studio. http://shiny.rstudio.com/

---

*bijuwang@uconn.edu