

Group Projects on bootstrapping

P8160 Advanced Statistical Computing

Group assignment is the same as in the Project 2

Requirements of a written report and in-class presentation:

On Apr 17th (Friday, tentative date), each group will give a presentation to demo your work and your findings from projects 2 and 3. Submit your written report with your R-codes afterward your presentation.

Protein Expression Data for Down syndrome

The data *Down.csv* consists of the expression levels of 77 proteins/protein modifications that produced detectable signals in the nuclear fraction of cortex. It has 1080 rows and 79 columns. The first column **MouseID** identifies individual mice; The column **2-78** are values of expression levels of 77 proteins. Column 79 indicates whether the mouse is a control or has Down syndrome.

The goal is to develop classification model based on the proteins expression levels.

Your to-do-list is

1. Apply the logistic-LASSO algorithm your group developed in Project 2 to select features for predicting Down syndrome.
2. Use the bootstrap-smoothing approach proposed in Efron (2014, JASA) to predict Down syndrome.
3. Compare the two approaches in predicting Down syndrome, and report your findings.
4. The researchers want to identify a subset of proteins that are significantly associated with the Down syndrome. How would you select the features in the bootstrap-smoothing approach? Explain the statistical justifications of your suggestion, implement them in R, and report your findings.

Answer: your answer starts here...