

Assignment1_project

Kishan Sarpangala

March 24, 2018

Question :

Project 1 n-fold cross-validation

This assignment builds upon the R/RStudio class and expands the n-fold cross-validation example.

Exercise 1.

For the assignment use the second dataset her2_lumA.txt.

Exercise 2.

Compute cross-validation estimates of accuracies for first 50 genes vs. PAM50 genes vs. all genes for both 3 - and 10 -fold cross-validation (3x2 table).

Exercise 3.

Create a R markdown document to report your result in a table format.

Exercise 4.

Comment on statistical significance of differences for different gene selection and n-fold cross-validation.

Exercise 5.

For up to 5 extra points replace current classifier with a logistic regression-based classifier and compare

Result with the simple centroid based.

Note:

The assignment is due date at the end of the spring break Ã¢â€œ March 28, 2018 midnight.

The submission should be zip compressed file named Ã¢â€œproject1-[your last name].zip which includes project1.Rmd and any supporting R files.

The zip file should be uploaded canopy. The assignment entry in canopy will be created shortly.