Project 1 – n-fold cross-validation

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

1. for the assignment use the second dataset her2_lumA.txt.
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).
3. create a R markdown document to report your result in a table format.
4. comment on statistical significance of differences for different gene selection and n-fold cross-validation.
5. for up to 5 extra points replace current classifier with a logistic regression-based classifier and compare result with the simple centroid based.

The assignment is due date at the end of the spring break – March 18, 2018 midnight.

The submission should be zip compressed file named "project1-[*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.