Statistical Analysis of Genetic Data
Math 155, Spring 2008
Jo Hardin
Project 10: PAM classification

Due Wednesday May 7

1 Instructions

1. The goal for this part of the project is to classify the arrays from your data using Prediction Analysis for Microarrays (PAM).

2. Like we did with the differential expression, you should break up your arrays into at least two different sub-groupings (e.g., in class I broke down my arrays first into 1 generation, 8 generations, 12 generations, and 18 generations; then I broke down my arrays into young and old).

2 Things to put on your web site (next)

- The classification done by PAM is supervised yet cross validated. That means that each test group is not part of the training model. I don't expect your classifications to be perfect!! You should try different options (sub-groupings, values of delta,...) to try to find the optimal classification.

- The PAM output has various plots and tables. You should put them on your web site (note: you can do cntrl-c to copy the plot and cntrl-v to paste the plot onto a word document). All plots should be well described.

- Don't try to put everything you do onto the web site. Put the relevant plots (e.g., two plots which compare results from two different sub-grouping).

- Make sure you explain any plot you put up. These aren't standard plots, so you'll really need to define the x- and y-axes as well as saying what the plotting symbols are.

- Give some insight to / thoughts about the results. For example, in my data, it was really the 18 generation arrays that were totally different from the others. Additionally, to get the "best" classification, I needed to keep thousands of genes! (Which seems untenable from a biologist's perspective.)