Title: Finding differences between two multivariate samples

Tarn Duong
School of Mathematics and Statistics,
University of New South Wales, Sydney, Australia

Abstract:

A common question in flow cytometry data analysis is where two multivariate samples are different. Differences between the samples can indicate differences in biological responses. Early attempts to answer this question are based on probability binning (a binary division algorithm) combined with classical chi-squared tests. Subsequently these have been found to be sub-optimal for flow cytometry data which typically have moderate dimension (around 3 to 20) and large sample sizes (in the order of 10 000 and 100 000).

Our proposal is to use the Patient Rule Induction Method (PRIM), a bump-hunting algorithm, combined with generalized chi-squared tests. We believe that our proposed method is more appropriate since PRIM is suited to moderate dimensions and generalized chi-squared tests are suited to large sample sizes. We briefly outline our theoretical reasoning and follow it with some case studies.