Disease Classification Using Microarray Data With Markov Blanket Filtering

Developments in mass-throughput molecular biology techniques over the past decade give us access to a vast amount of high-dimensional microarray data which display tens of thousands of genes (features) simultaneously at a molecular level with a very small number of samples. For data of such high dimension, it is crucial to reduce their dimensions before performing downstream analysis such as prediction and classification of relevant diseases. In this paper, we summarize some popular dimension reduction methods and evaluate the performance of a promising method, which reduces data dimensions by finding a Markov blanket around the target variable, in disease classification using real microarray data.

Yingcheng Luo, MS candidate

stat.uchicago.edu