

Following the work of Dai et al. (2006), Tan et al. (2008) and Osareh and Shadgar (2009) we examine the performance of several clustering and classification algorithms, together with common dimension reduction methods, when applied to two very different microarray datasets (yeastCC and Breast Cancer).

Clustering techniques examined include two main groups - partitioning and hierarchical methods as well as Plaid Models (Lazzeroni and Owen, 2002). This study examined the partitioning methods k-means, k-medoids, and Self-Organizing Maps (SOMs), and a hierarchical method - agglomerative hierarchical with mean linkage. Plaid models, which were specifically designed to accommodate the characteristics of microarray data, were also studied.

Classification methods adaBoost, C4.5, single classification trees (both *rpart* and *tree* library from R), neural networks, support vector machines, and random forest were used in combination with two dimension reduction techniques: partial least squares and principal components analysis. The suitability of these methods in combination with dimension reduction techniques was assessed using experimental design.