

## **Support Vector Clustering of Microarray Data**

*Ozlem Yilmaz, Luke.E.K.Achenie, Ranjan Srivastava*

*Department of Chemical Engineering, University of Connecticut, Storrs, CT, USA*

Microarray experiments enable us to look at expression profiles of tens of thousands of genes at the same time. Clustering is one method that can be applied to the data so that we can determine useful information hidden in large sets such as finding co-regulated and anti-regulated genes.

Support vector machines (SVMs) are powerful techniques that have been used not only for classification, but also for clustering purposes. Our goal is to cluster biological data, primarily microarray data, using a novel support vector clustering (SVC) algorithm based on SVM. The SVC employs kernel functions that have tunable parameters. These parameters, together with the soft margin constant, are usually set arbitrarily and therefore sub-optimally. To avoid this, we have developed an optimization strategy which utilizes the SVC. A near-optimal set of parameters is determined using simulated annealing, which is a stochastic global optimization algorithm. Using these parameters, our goal is to develop a consistent and biologically relevant technique for clustering.

Clustering results using SVC are reported for three data sets and compared to those obtained using traditional techniques. The first data set tested is an iris data set from the University of California-Irvine (UCI) repository. The data analyzed consists of three classes of a type of iris plant distinguished by sepal and petal measurements. The second system analyzed (thyroid gland data) is also from UCI and contains three classes. Each class corresponds to diagnosis of a thyroid gland function based on laboratory tests. Finally, the SVC algorithm is applied to yeast cell cycle data and the results obtained are compared with results from open literature.