SURVEY ON CLUSTER ANALYSIS FOR GENE EXPRESSION DATA

BEENA BAKSHI AND U. S PANDEY

School of Open Learning. Delhi University, India.

Abstract

Cluster analysis seeks to partition a given data set into groups based on specified features so that the data points within a group are more similar to each other than the points in different groups. These clustering algorithms have been proven useful for identifying biologically relevant groups of genes and samples. In this paper, we briefly introduce the concepts of microarray technology and discuss the basic elements of clustering on gene expression data. Clustering is a problem which finds application everywhere in data analysis. Traditional Clustering problems try to optimize the distances from the center/mean of the cluster. They do not take the connectedness of the points under consideration. Connected k center and connected k means are the problems in which one requires that given some relationship data amongst the objects, the objects within a cluster are related.