

CC Biclustering analysis of gene expression data with *biclust* R package

Daianna González Padilla

23/2/2022

Keywords

Biclustering, biclust, CC algorithm, gene expression.

Abstract

This work accomplishes a CC biclustering analysis to exemplify its functionality and execution in R programming language with real expression data from differentially expressed genes using biclust package. It also compares the outcome with clusters obtained by hierarchical clustering methods and with the clustering from a differential gene expression analysis previously done [0]. .

Introduction

Biclustering, also known as co-clustering, two-dimensional clustering and two-way clustering, is an unsupervised data mining technique that performs clustering on the rows and columns of a data matrix at the same time considering only a subset of relevant features when grouping objects into clusters, however, not only the objects are clustered but also their features [1]. The goal of this clustering algorithm is to find sub-matrices in the dataset (*e.g.* subsets of objects and attributes) whose elements exhibit significant homogeneity or in other words, that are as similar as possible to each other and as different as possible to the rest [2,8]. Such submatrices consist of subsets of columns that in turn determine row assignment [3].