Rapport package team

Hierarchical Cluster Analysis

2011-04-26 20:25 CET

## Description

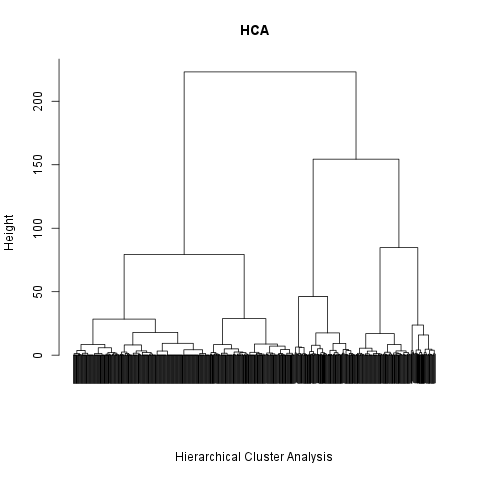
In this template Rapporter will present you Hierarchical Cluster Analysis.

### Introduction

[Hierarchical Cluster Analysis](http://en.wikipedia.org/wiki/Hierarchical_clustering) is a data mining method which seeks to build a hierarchy of clusters. Clusters are calculated based on the distances between the observations. At the beginning each observation is assigned to be a single cluster, later in every round the most similar clusters will be joined until all observations are in one cluster. One should not mix it up with [K-means Cluster Analysis](http://en.wikipedia.org/wiki/K-means_clustering), which calculates the clusters based on the final numbers of them.

### HCA

Below you can see on the plot how the clusters were made, how the observations were paired with each other. The horizontal linkage between the vertical lines indicates the stage where two clusters joined to each other. In the bottom of the plot you can see the clustering process in an other way, for each observations the shorter lines indicate later clustering.

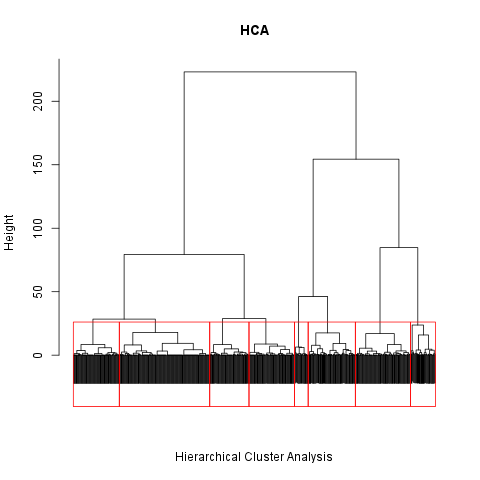
[](plots/HierarchicalClusterAnalysis.tpl-1-hires.png)

We can say that *438* observations have the same values on the used variables, so they were joined in the first *438* round. After that *40* times there were only made clusters with 2 observations, the first cluster that contains 3 was made in the round *478*.

##### Optimal number of clusters

According to the BIC for EM initialized by hierarchical clustering for parameterized Gaussian mixture models, the optimum numbers of the clusters are *8*.

Let's see how the Dendogram looks like when we the optimal number of the clusters (*8*) plotted in it.

[](plots/HierarchicalClusterAnalysis.tpl-2-hires.png)

## Description

In this template Rapporter will present you Hierarchical Cluster Analysis.

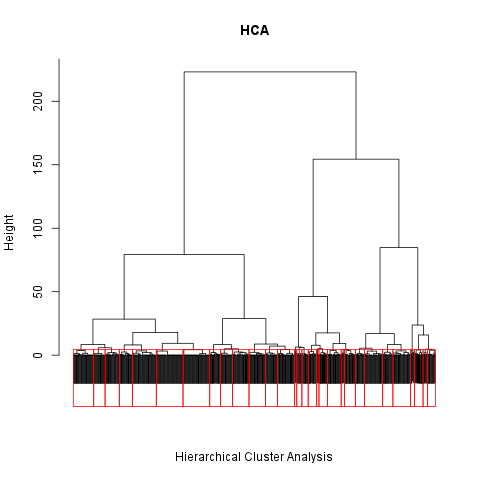
### Introduction

[Hierarchical Cluster Analysis](http://en.wikipedia.org/wiki/Hierarchical_clustering) is a data mining method which seeks to build a hierarchy of clusters. Clusters are calculated based on the distances between the observations. At the beginning each observation is assigned to be a single cluster, later in every round the most similar clusters will be joined until all observations are in one cluster. One should not mix it up with [K-means Cluster Analysis](http://en.wikipedia.org/wiki/K-means_clustering), which calculates the clusters based on the final numbers of them.

### HCA

Below you can see on the plot how the clusters were made, how the observations were paired with each other. The horizontal linkage between the vertical lines indicates the stage where two clusters joined to each other. In the bottom of the plot you can see the clustering process in an other way, for each observations the shorter lines indicate later clustering.

The red boxes shows the last 30 clusters.

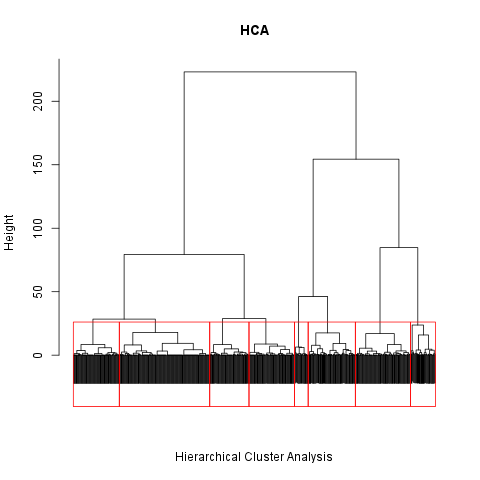
[](plots/HierarchicalClusterAnalysis.tpl-3-hires.png)

We can say that *438* observations have the same values on the used variables, so they were joined in the first *438* round. After that *40* times there were only made clusters with 2 observations, the first cluster that contains 3 was made in the round *478*.

##### Optimal number of clusters

According to the BIC for EM initialized by hierarchical clustering for parameterized Gaussian mixture models, the optimum numbers of the clusters are *8*.

Let's see how the Dendogram looks like when we the optimal number of the clusters plotted in it.

[](plots/HierarchicalClusterAnalysis.tpl-2-hires.png)

## Description

In this template Rapporter will present you Hierarchical Cluster Analysis.

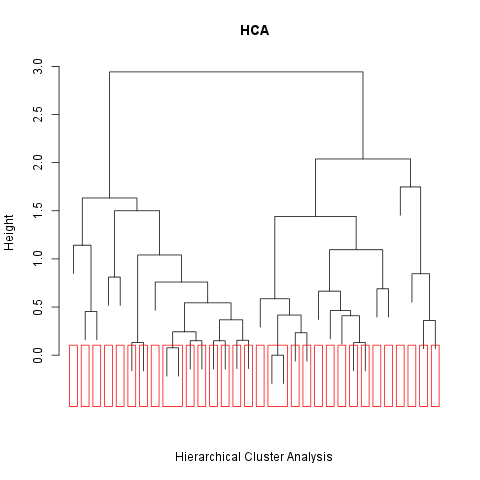
### Introduction

[Hierarchical Cluster Analysis](http://en.wikipedia.org/wiki/Hierarchical_clustering) is a data mining method which seeks to build a hierarchy of clusters. Clusters are calculated based on the distances between the observations. At the beginning each observation is assigned to be a single cluster, later in every round the most similar clusters will be joined until all observations are in one cluster. One should not mix it up with [K-means Cluster Analysis](http://en.wikipedia.org/wiki/K-means_clustering), which calculates the clusters based on the final numbers of them.

### HCA

Below you can see on the plot how the clusters were made, how the observations were paired with each other. The horizontal linkage between the vertical lines indicates the stage where two clusters joined to each other. In the bottom of the plot you can see the clustering process in an other way, for each observations the shorter lines indicate later clustering.

The red boxes shows the last 30 clusters.

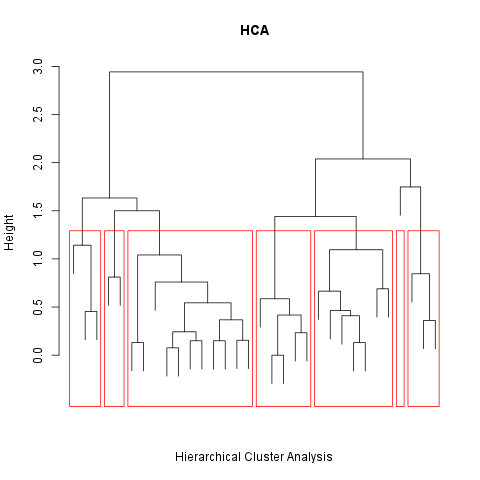
[](plots/HierarchicalClusterAnalysis.tpl-4-hires.png)

We can say that *1* observations have the same values on the used variables, so they were joined in the first *1* round. After that *8* times there were only made clusters with 2 observations, the first cluster that contains 3 was made in the round *9*.

##### Optimal number of clusters

According to the BIC for EM initialized by hierarchical clustering for parameterized Gaussian mixture models, the optimum numbers of the clusters are *7*.

Let's see how the Dendogram looks like when we the optimal number of the clusters plotted in it.

[](plots/HierarchicalClusterAnalysis.tpl-5-hires.png)

This report was generated with [R](http://www.r-project.org/) (3.0.1) and [rapport](http://rapport-package.info/) (0.51) in *4.12* sec on x86\_64-unknown-linux-gnu platform.

