Introduction to hierarchical clustering analysis

As indicated by the term ‘hierarchical’, the method seeks to build clusters based on hierarchy. Generally, there are two types of clustering strategies: Agglomerative and Divisive. Here, we mainly focus on the agglomerative approach, which can be easily pictured as a ‘bottom-up’ algorithm. Observations are treated separately as singleton clusters. Then, compute the Euclidean distance of each pair and successively merge the most similar clusters. Repeated the previous step until the final optimal clusters are formed.

There are four methods for combining clusters in agglomerative approach. The one we choose to use is called Ward’s Method. Unlike the others. Instead of measuring the distance directly, it analyzes the variance of clusters. Ward’s is said to be the most suitable method for quantitative variables.

Ward’s starts with *n* singleton clusters Like other agglomerative clustering methods. Then, the pre-defined clusters combine as one, containing all objects. The goal of each step of the computation is to minimize the [variance](https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/variance/) of new cluster. Such measurement is called the sum of squares index.

To calculate the index, first, we need to find the mean of each cluster, then calculate the distance between each point within one cluster. The result at this point is called the cluster’s mean. Next, compute the difference of distances, square and sum them up. Last, we follow the similar idea as cluster’s mean to define the total sum of squares (e.g., TSS) but compare against the grand mean of each variable.

Let Xijk denote the value for variable k in observation j belonging to cluster i.

**Error Sum of Squares**:

**Total Sum of Squares**:

**R-Square**: r2=(TSS-ESS)/TSS

Therefore, our goal of minimize the variance can be seen equally as maximize the r^2 value.

Reference

<https://en.wikipedia.org/wiki/Hierarchical_clustering#Divisive_clustering>

<https://onlinecourses.science.psu.edu/stat505/node/146/>

<https://nlp.stanford.edu/IR-book/html/htmledition/hierarchical-agglomerative-clustering-1.html>