Assignment Instructions: Assignment 5

Purpose

The purpose of this assignment is to use Hierarchical Clustering. The assignment requires you to:

- Construct a hierarchical clustering model in R and to comment on advantages of that model compared to K-means algorithm.

Directions

The dataset Cereals.csv includes nutritional information, store display, and consumer ratings for 77 breakfast cereals.

Data Preprocessing. Remove all cereals with missing values.

- Apply hierarchical clustering to the data using Euclidean distance to the normalized measurements. Use Agnes to compare the clustering from single linkage, complete linkage, average linkage, and Ward. Choose the best method.
- Comment on differences between hierarchical Clustering and K-means.
- How many clusters would you choose?
- Comment on the structure of the clusters and on their stability. Hint: To check stability, partition the data and see how well clusters formed based on one part apply to the other part. To do this:
    - Cluster partition A
    - Use the cluster centroids from A to assign each record in partition B (each record is assigned to the cluster with the closest centroid).
    - Assess how consistent the cluster assignments are compared to the assignments based on all the data.
- The elementary public schools would like to choose a set of cereals to include in their daily cafeterias. Every day a different cereal is offered, but all cereals should support a healthy diet. For this goal, you are requested to find a cluster of "healthy cereals." Should the data be normalized? If not, how should they be used in the cluster analysis?
- How do you compare hierarchical clustering and k-means? What are they main advantages of hierarchical clustering compared to k-means?

File Attached: Cereals.csv

Module Learning Outcomes

- Identify the advantages of deploying a hierarchical clustering algorithm
- Compare hierarchical clustering algorithms against k-means and DBSCAN methods
- Implement a hierarchical clustering model in R and interpret the output of the model

Requirements

All due dates are included in the Assignment Schedule.

General Submission Instructions

All work must be your own. Copying other people's work or from the Internet is a form of plagiarism and will be prosecuted as such.

- Upload a pdf document to your git repository. Name your file Username_#.ext, where Username is your Kent State User ID (the part before @), and # is the Assignment number. In this case, 5.

Provide the link to your git repository in Blackboard Learn for the assignment.