ENGINEERING DIVISION | NYU ABU DHABI

ENGR-UH 3332
Applied Machine Learning

Bonus Project – Hierarchical clustering

Due Date: Refer to NYU Class

BONUS (30 points)

Hierarchical clustering involves creating clusters that have a predetermined ordering. For example, all files and folders on the hard disk are organized in a hierarchy. More details, please refer to lecture slides.

Dataset

In this bonus project you will work on a given Mall Customer dataset (Mall_Customers.csv)

Requirements

1. Implement a hierarchical clustering model using 'Ward' distance matrix for this dendrogram.

   Ward distance matrix

   We will use "ward" distance matrix for this dendogram.

$$d(u, v) = \sqrt{\frac{|v| + |s|}{T} d(v, s)^2 + \frac{|v| + |t|}{T} d(v, t)^2 - \frac{|v|}{T} d(s, t)^2}$$

   Where u is the newly joined cluster consisting of clusters s and t,v is unused cluster in the forest, T= $|v| + |s| + |t|$ and $|*|$ is the cardinality of its argument. This is also known as the incremental algorithm.

2. Plot the clusters and label customer types

Deliverables

A .ipynb file containing the following:

1. Source code
2. Detailed description of the project if needed

Before submitting your project, please make sure to test your program on the given dataset.

Notes

You may discuss the general concepts in this project with other students, but you must implement the program on your own. ***No sharing of code or report is allowed.*** *Violation of this policy can result in a grade penalty.*

Late submission is acceptable with the following penalty policy:
10 points deduction for every day after the deadline