CS 663 - Machine Learning Spring, 2020

Lab 04 - Dendrogram

As seen during class, hierarchical clustering is one clustering algorithm. It is especially useful for producing a dendrogram — a view of the data in which data points are connected to a binary tree. In this lab, your task is to recreate the visual work in class, show how you are able to apply the technique to a previously-unseen dataset and recreate a dendrogram by following reference documentation.

Data

There are two datasets for this lab:

- The (reduced) US Census data
- The NIPS Conference Papers dataset

Process

Your implementation must:

- Use the starting data and fill in your name and USF email in the readme
- Read the data from and apply hierarchical clustering on the two datasets above to produce dendrograms
- Use a Jupyter notebook or Python script to complete your implementation.

Grading

Grades for this assignment will be determined by the grader as follows:

- 100% = Code functions, is well-documented and clearly shows the relationship of data points in a binary tree.
- 75% = Code functions but is not well-documented or does not clearly show the relationship of data points in a binary tree.
- 50% = Code functions but is not well-documented -AND- does not clearly show the relationship of data points in a binary tree.
- 0% = No submission / code does not function.

Submission

Submit your code to Github, and submit the link to your repo on Canvas. Late submissions will be penalized or not accepted.

The deadline is 11:59 PM on February 17, 2020.