Guided Projects Artificial Intelligence & Machine Learning

Guided Projects: Unsupervised Learning

Hierarchical K-Means: Construction of Hashing Tree

Hierarchical clustering is the hierarchical decomposition of the data based on **group similarities**. It allows us to build tree structures from data similarities and see how different sub-clusters relate to each other, and how far apart data points are. It gives us a tree-type structure based on the hierarchical series of nested clusters. **A diagram called Dendrogram graphically represents this hierarchy** and is an inverted tree that describes the order in which factors are **merged, or clusters are broken apart**.

Question:

Perform **Hierarchical Clustering** from scratch and also **using sklearn** to perform wholesale customer segmentation based on their **annual spending on products**. You can use **this dataset**.
Use the threshold to
1. Divide the dataset into **two clusters**.
2. To divide the dataset into **k clusters**, such that the distance between the **two clusters is greater** than a given threshold (this **threshold** can be anything passed to the function).

Dataset Link: Wholesale customers data

https://archive.ics.uci.edu/ml/machine-learning-databases/00292/Wholesale%20customers%20data.csv