

CAPSTONE PROJECT REPORT

Hierarchical K-Means: Construction of Hashing Tree

Name - Paleti Samuel Yashaswi
Course - Machine learning and AI
Duration - 24 months
Question - 8

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).

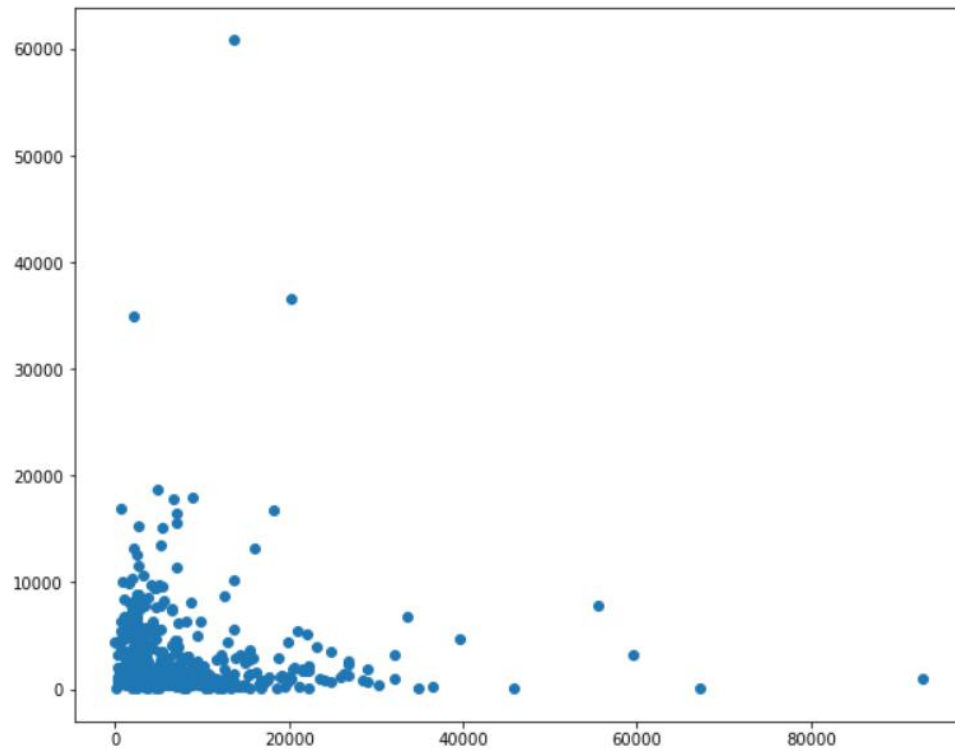
Dataset Link: Wholesale customers data
<https://archive.ics.uci.edu/ml/machine-learningdatabases/00292/Wholesale%20customers%20data.csv>

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
%matplotlib inline

customer_data = pd.read_csv('Wholesale customers data.csv')
data = customer_data.iloc[:, 4:6].values

labels = range(1, 11)
plt.figure(figsize=(10, 8))
plt.subplots_adjust(bottom=0.1)
plt.scatter(data[:,0],data[:,1], label='True Position')
plt.xlabel('milk')
plt.ylabel('grocery')

plt.show()
```



```
import scipy.cluster.hierarchy as shc

plt.figure(figsize=(10, 7))
plt.title("Customer Dendograms")
dend = shc.dendrogram(shc.linkage(data, method='ward'))
```

