hierarchical-clustering-1

August 28, 2023

#Name : Algam Sai Krishna #Roll no: 21X05A6703 #Branch : Data Science #College: Narsimha Reddy Engineering College

#Project Title Analysis and prediction of "Mall_Customers" of american mall market called PHONIX Mall.To find out on the basics requirements of dendrogram using scipy graphics library with the help of "scipy.cluster.hierarchy",to ace the no_of linkage of a clustering to predict.

#Problem Statement The American finance market clients as per the rate of GDP of 2011 found has hieghest number growth in thier bussiness market.

As the data science engineer find out which hierarcy cluster gives maximum linkage in upcoming future

#Task1: import the libery and datasets #Task2:Using the dendogram to find the optimal number of clusters #Task3:Create a hierarchy model and visualize the cluster with the help of matplotlib

1 Hierarchical Clustering

1.1 Importing the libraries

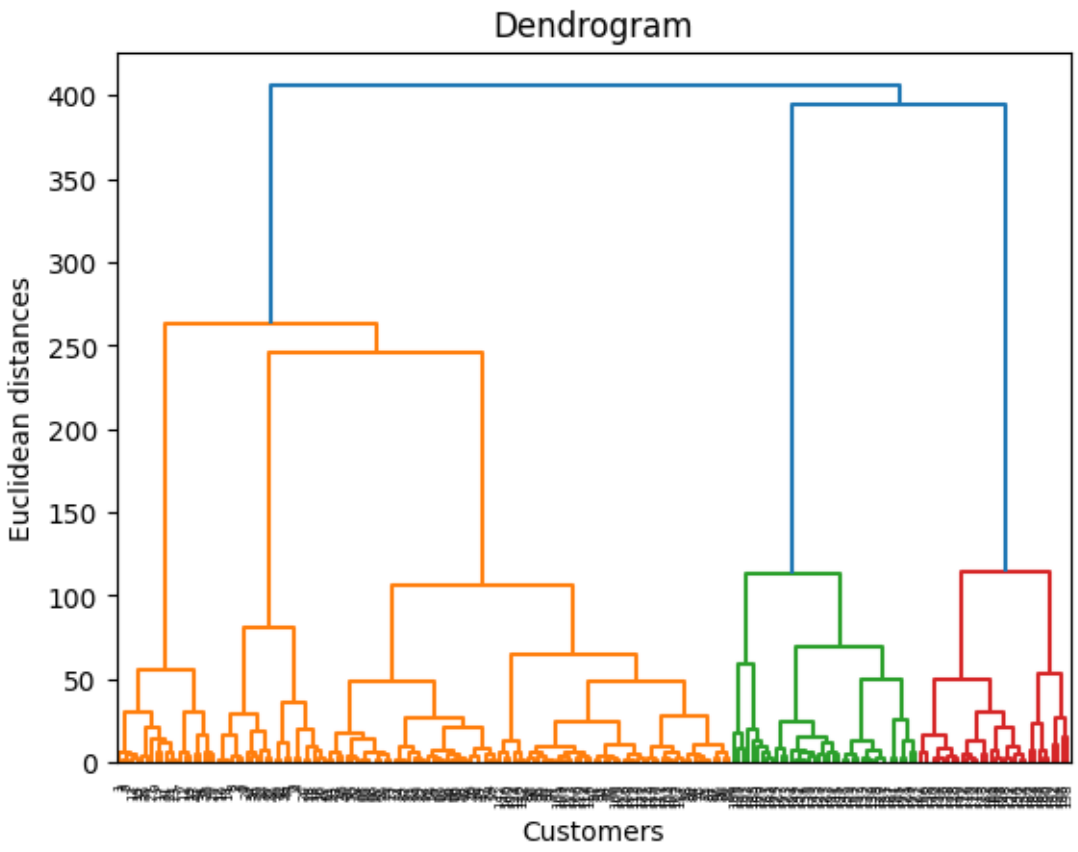
[4]:
```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

1.2 Importing the dataset

[6]:
```
dataset = pd.read_csv('Mall_Customers.csv')
X = dataset.iloc[:, [3, 4]].values
```

1.3 Using the dendrogram to find the optimal number of clusters

[7]:
```
import scipy.cluster.hierarchy as sch
dendrogram = sch.dendrogram(sch.linkage(X, method = 'ward'))
plt.title('Dendrogram')
plt.xlabel('Customers')
plt.ylabel('Euclidean distances')
plt.show()
```



1.4 Training the Hierarchical Clustering model on the dataset

[8]:
```
from sklearn.cluster import AgglomerativeClustering
hc = AgglomerativeClustering(n_clusters = 5, affinity = 'euclidean', linkage =␣
 ↪'ward')
y_hc = hc.fit_predict(X)
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_agglomerative.py:983:
FutureWarning: Attribute `affinity` was deprecated in version 1.2 and will be
removed in 1.4. Use `metric` instead
  warnings.warn(
```

1.5 Visualising the clusters

[9]:
```
plt.scatter(X[y_hc == 0, 0], X[y_hc == 0, 1], s = 100, c = 'red', label =␣
 ↪'Cluster 1')
plt.scatter(X[y_hc == 1, 0], X[y_hc == 1, 1], s = 100, c = 'blue', label =␣
 ↪'Cluster 2')
```

```
plt.scatter(X[y_hc == 2, 0], X[y_hc == 2, 1], s = 100, c = 'green', label =␣
 ↪'Cluster 3')
plt.scatter(X[y_hc == 3, 0], X[y_hc == 3, 1], s = 100, c = 'cyan', label =␣
 ↪'Cluster 4')
plt.scatter(X[y_hc == 4, 0], X[y_hc == 4, 1], s = 100, c = 'magenta', label =␣
 ↪'Cluster 5')
plt.title('Clusters of customers')
plt.xlabel('Annual Income (k$)')
plt.ylabel('Spending Score (1-100)')
plt.legend()
plt.show()
```

#Conclusion According to the model building as a engineer my prediction is cluster number 3 as give highest number of linkage.