hierarchical-clustering-2

August 28, 2023

NAME : D.SHYAM PRASAD

ROLL NO : 21X05A6714

BRANCH : CSE(DATA SCIENCE)

COLLEGE : NRCM

PROJECT TITLE:

Analysis and predection of Malls customer.cs file of american Mall market called as phonex mall, find out on basis of client requriments of dendograms using scipy graphics library with the help of "scipy cluster. hierarchy to ace the number of linkage of clustering to predict

PROBLEM STATEMENT:

The american finance market clients as per the rate od GDP of 2011 found as highest number of growth in there business market. As a datascience engineer find out which hierarchy cluster give maximum linkage in upcoming future

Task-1: With help of sipcy library import the library and import datasets.

Task-2: Using the dendrogram to find the optimal number of clusters.

Task-3: Create a hirerachy model and viuliaze the cluster with help of matplot library.

1 Hierarchical Clustering

1.1 Importing the libraries

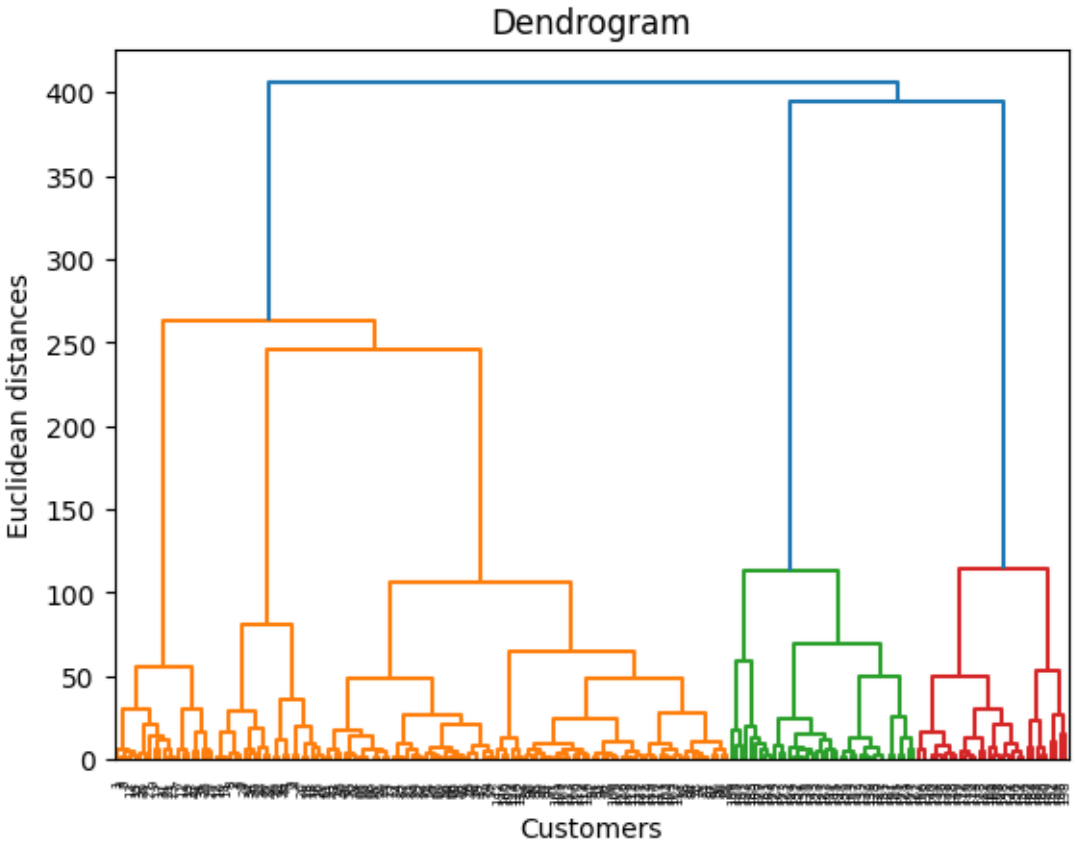
[ ]:
```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

1.2 Importing the dataset

[ ]:
```
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

```
[ ]: 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()
```



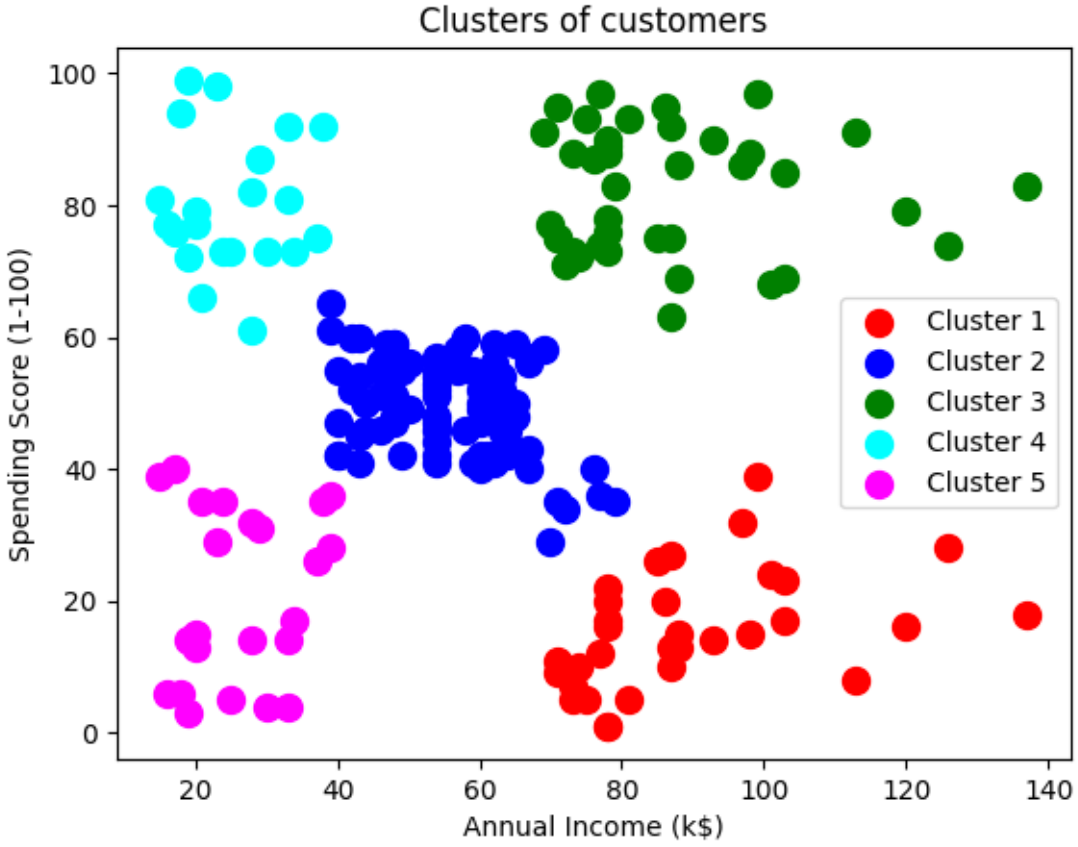
1.4 Training the Hierarchical Clustering model on the dataset

```
[ ]: 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

[13]:
```
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 model buliding as a engineer my prediction is cluster number has give highest number

of linkage.

INSIGHTS:

Cluster-1 : contains {red} which shows that unsupervised learning cluster has maximum ucliding distance from the centroid up to annual income approximate 139ks.

Cluster-2: contains {blue} which shows that unsupervised learning cluster as maximum ucliding distance from centroid up to annual income approximately 79 t0 80ks.

Cluster-3: contains {green} which shows that unsupervised learning cluster as maximum ucliding distance from centroid up to annual income approximately 139ks.

Cluster-4: contains {cyan} color which shows that unsupervised learning cluster has maximum ucliding distance from the centriod upto annual income appropriate 140ks.

Cluster-5: contains {magenta} color which shows that unsupervised learning cluster has maximum ucliding distance from the centriod upto annual income appropriate 41ks