hierarchical-clustering-1-1

August 29, 2023

#Name:Baddam Poojitha #Roll No:21X05A6707 #Branch:Data Science #College:Narasimha Reddy Engineering College

#Project Title: Analysis and prediction of “mall_customers_csv” of American mall market called as Phonix mall,find out on the basis of clients requirements of dendrograms using scipy graphics library with the help of “scipy.cluster.hierarchy”,to ace the no of linkage of the clustering to predict

#Problem Statement: The American Finance market clients as per the rate for the GDP of 2011 found as highest no of growth in their business market.

As Datascience enigineer find out which hierarchy cluster gives maximum limkage in upcoming future

#TASK: 1.Import the library and datasets

2.Using the dendogram to find the optimal No.of Clusters.

3.Create the hierarchy model and visualize the cluster with the help of matplot library

1 Hierarchical Clustering

1.1 Importing the libraries

[1]:

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

1.2 Importing the dataset

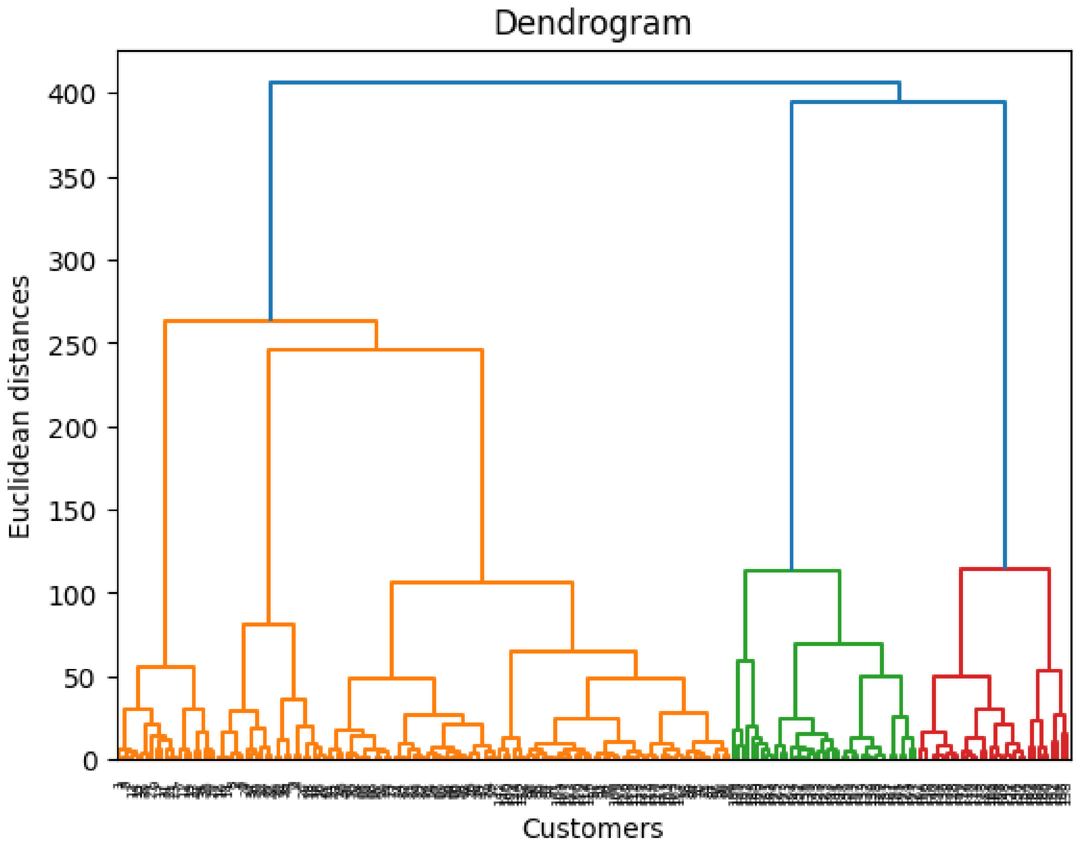
[4]:

```
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

[5]:

```
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()
```



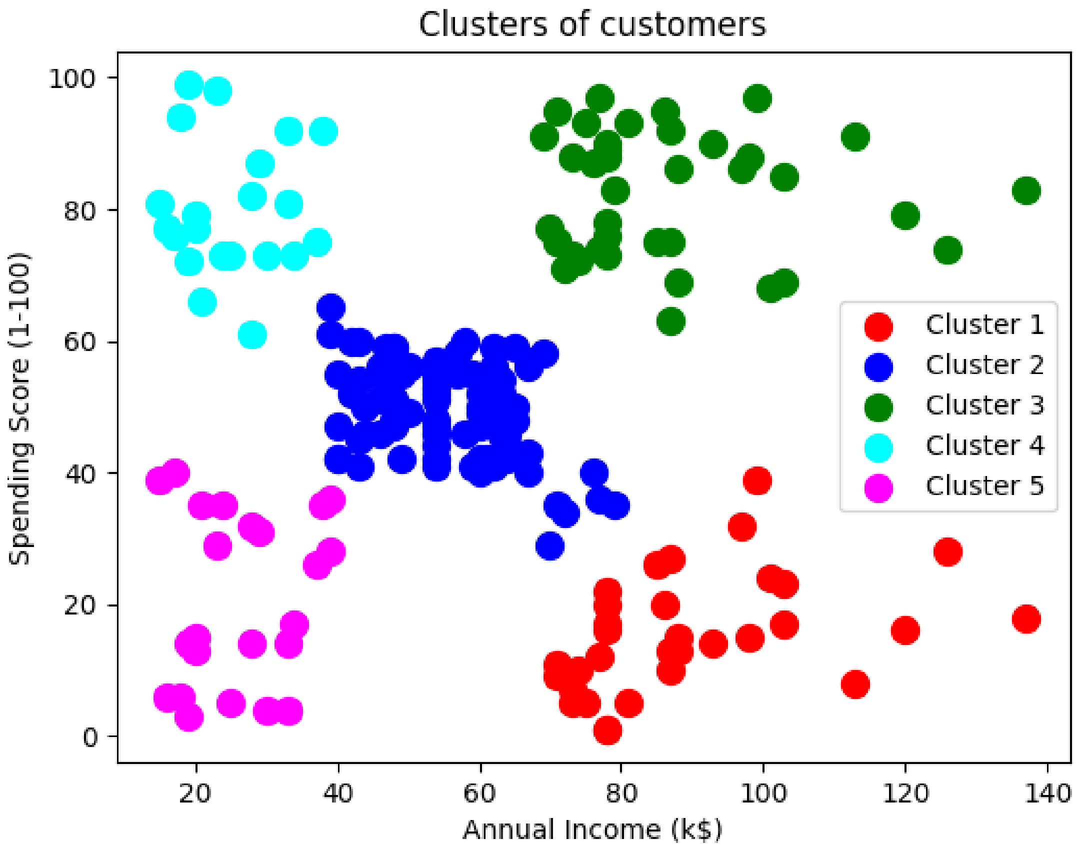
1.4 Training the Hierarchical Clustering model on the dataset

[6]:
```python
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

[7]:
```python
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 engineer my prediction is cluster NO.3 as give highest number of linkage.

#Insights: 1.Cluster 1 contains(Red) color which shows that unsupervised learning cluster has maximum euclidean distance from the centroid upto annual income approximate 139ks.

2.Cluster 2 contains(Blue) color which shows that unsupervised learning cluster has maximum euclidean distance from the centroid upto annual income approximate 79-80ks.

3.Cluster 3 contains(Green) color which shows that unsupervised learning cluster has maximum euclidean distance from the centroid upto annual income approximate 139ks.

4.Cluster 4 contains(Cyan) color which shows that unsupervised learning cluster has maximum euclidean distance from the centroid upto annual income approximate 39-40ks.

5.Cluster 5 contains(Pink) color which shows that unsupervised learning cluster has maximum euclidean distance from the centroid upto annual income approximate 40-41ks.