hierarchical-clustering2

August 28, 2023

##NAME : K.NIKHIL CHARY ##PIN-NO : 21X05A6726 ##BRANCH : DATA SCIENCE ##COLL : NRCM

1 Hierarchical Clustering

#project title: Analysis and prediction of "mall_customer.csv" of American mall market called as phonix mall, find out on basic of clients requirments of dendrogram using scipy graphic 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 of gdp of 2011 found as higest no of growth in the business market.

AS a data science engineerfind out which hierarchy clusteer give maximum linkage in uo comming future

###TASK 1 with the help of spicy library import the library and import datasets ###TASK 2 Using the dendrogram to find optimal no of clusters ###TASK 3 Create a hierarchy model and visualize the cluster with the help of matplot library

1.1 Importing the libraries

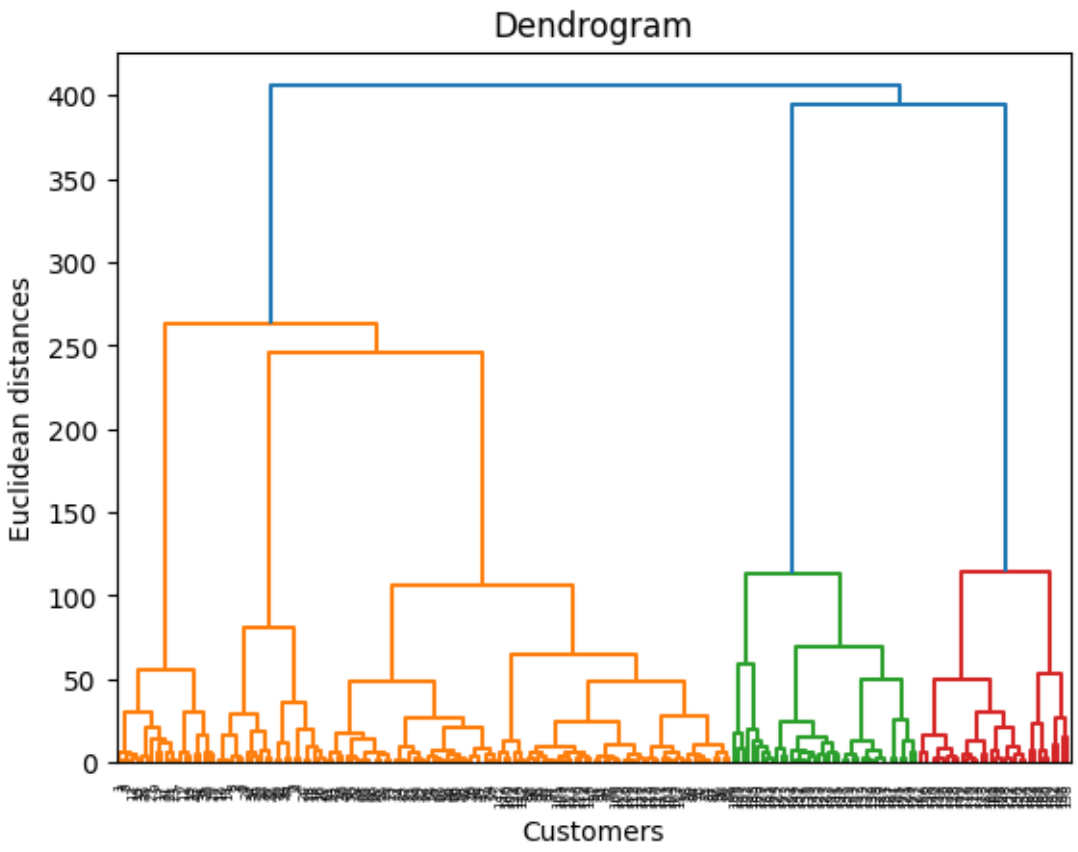
[ ]:
```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

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

[ ]:
```
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 may prediction is cluster-3 has give highest no of linkage

[ ]: