hierarchical-clustering

August 28, 2023

###Name : Boddupalli Renuka ###Class : 4th year ###Branch : Data science(cse) ###Roll no : 21X05A6709

##Project title: ###Analysis and prediction of small_custumers.csv file of American mall market called as phonix small,find out on the basis of clients requirements of Dendograms using scipy graphics librari withnthe 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 of GDP of 2011 found as highest no of growth in there business market. ###As a Datasciece Engineer find out which hirarchy cluster give maximum linkage in upcoming future.

##Task 1 ###With the help of scipy library import the libraries and import the Dataset ##Task 2 ###using the dendogram to find the optimal number of clusters ##Task 3 ###Create a hirarchical model and visualize the cluster with the help of matplotlib.

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()
```
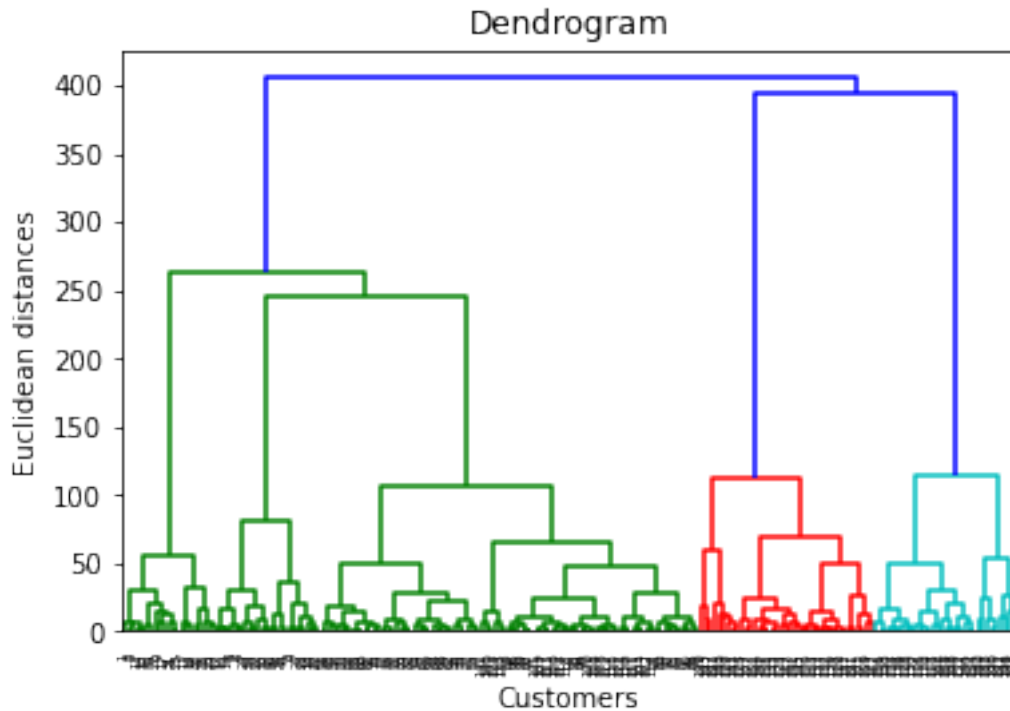
1.4 Training the Hierarchical Clustering model on the dataset

```python
from sklearn.cluster import AgglomerativeClustering
hc = AgglomerativeClustering(n_clusters = 5, affinity = 'euclidean', linkage =
↪'ward')
y_hc = hc.fit_predict(X)
```

1.5 Visualising the clusters

```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 a engineer my prediction is cluster no 3 as give higher no of linkage.