## → Practical 8:

## Agglomerative Hierarchical Clustering

## Submitted By:

- Aman Kumar
- · 2019UC01634
- CoE 2

```
# imports
import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
from sklearn.cluster import AgglomerativeClustering
from sklearn import datasets
import scipy.cluster.hierarchy as sch

dataset = pd.read_csv('Mall_Customers.csv')

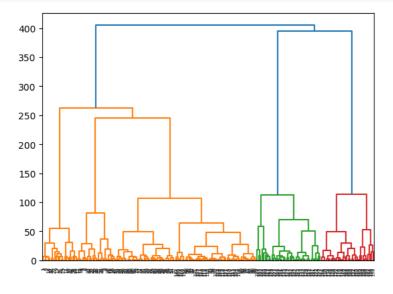
X = dataset.iloc[:, [3, 4]].values
```

dataset.head()

	CustomerID	Genre	Age	Annual Income (k\$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

We will be using Annual Income and Spending Score as the basis of our classification.

```
dendrogram = sch.dendrogram(sch.linkage(X, method='ward'))
```

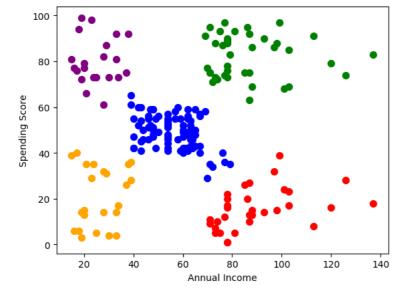


```
model = AgglomerativeClustering(n_clusters=5, metric='euclidean', linkage='ward')
model.fit(X)
labels = model.labels_
```

labels

```
1, 2, 0, 2, 0, 2, 0, 2, 0, 2, 1, 2, 0, 2, 1, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2, 0, 2])
```

```
plt.scatter(X[labels==0, 0], X[labels==0, 1], s=50, marker='o', color='red')
plt.scatter(X[labels==1, 0], X[labels==1, 1], s=50, marker='o', color='blue')
plt.scatter(X[labels==2, 0], X[labels==2, 1], s=50, marker='o', color='green')
plt.scatter(X[labels==3, 0], X[labels==3, 1], s=50, marker='o', color='purple')
plt.scatter(X[labels==4, 0], X[labels==4, 1], s=50, marker='o', color='orange')
plt.xlabel('Annual Income')
plt.ylabel('Spending Score')
plt.show()
```



Colab paid products - Cancel contracts here