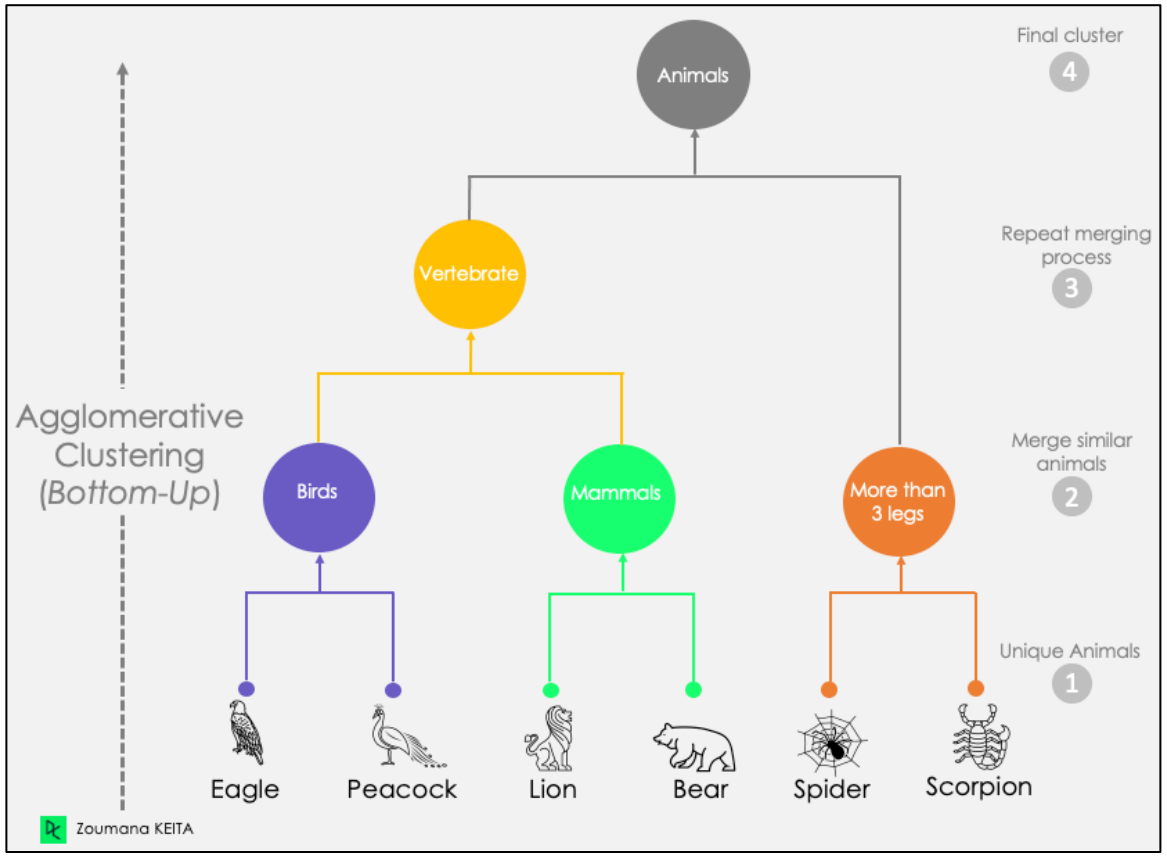
Hierarchical Clustering

UTKARSH GAIKWAD

CLASS STARTS SHARP AT 8:05 PM

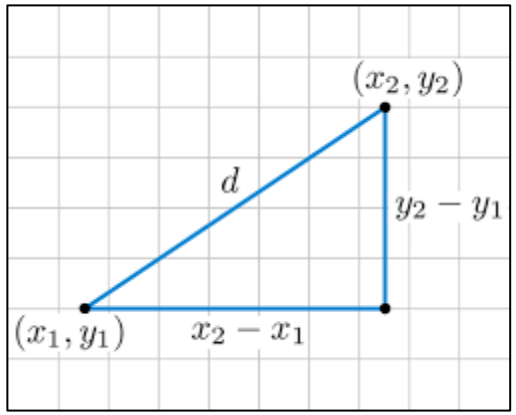
Hierarchical Clustering Explained

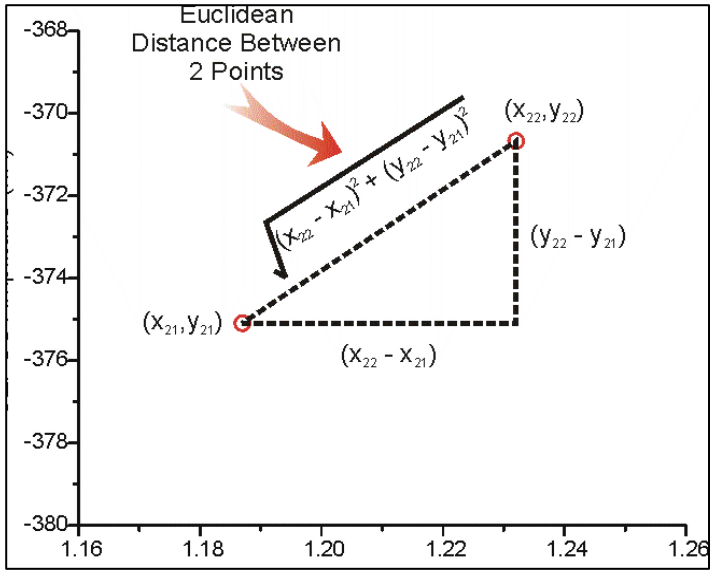


Divisive

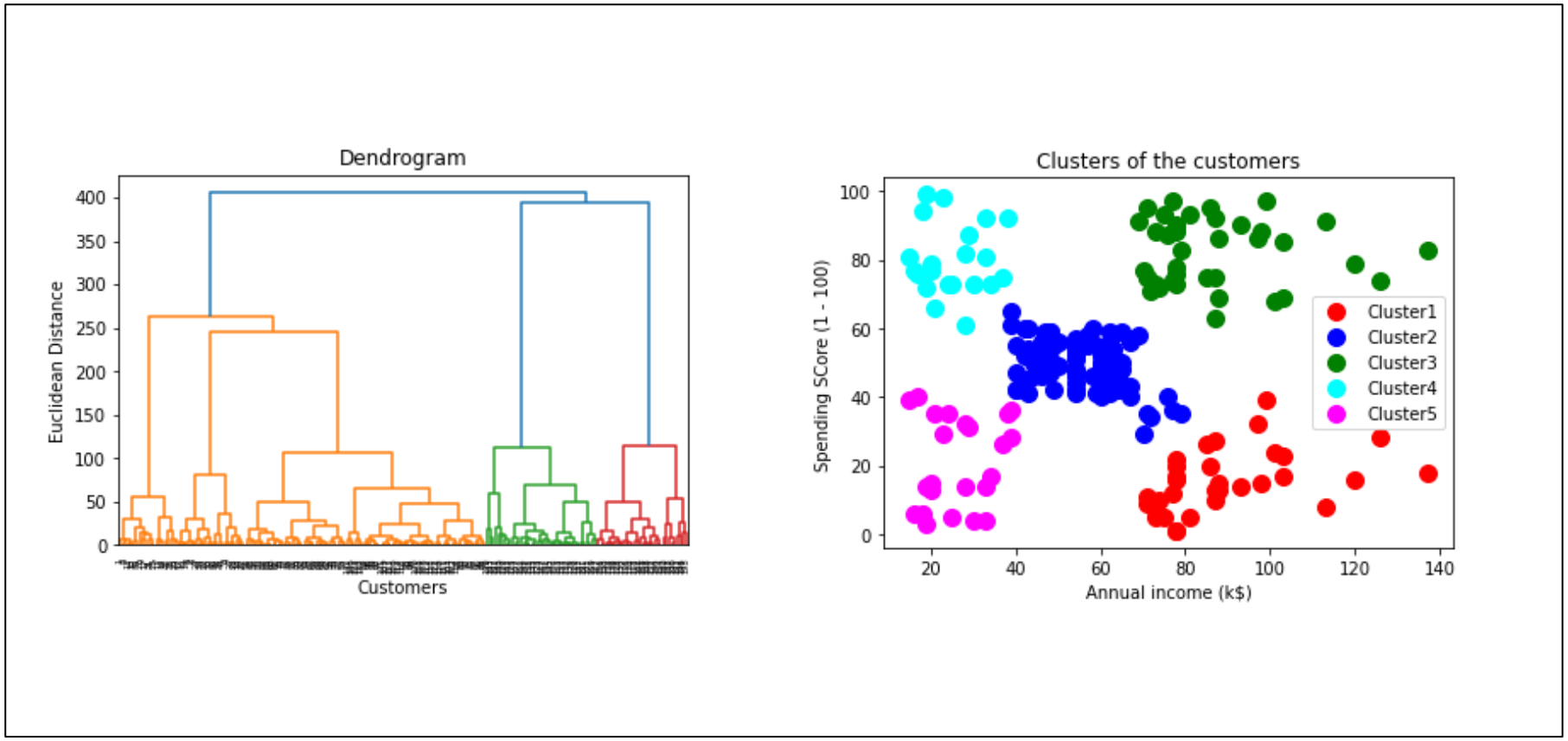
Euclidian Distance

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$





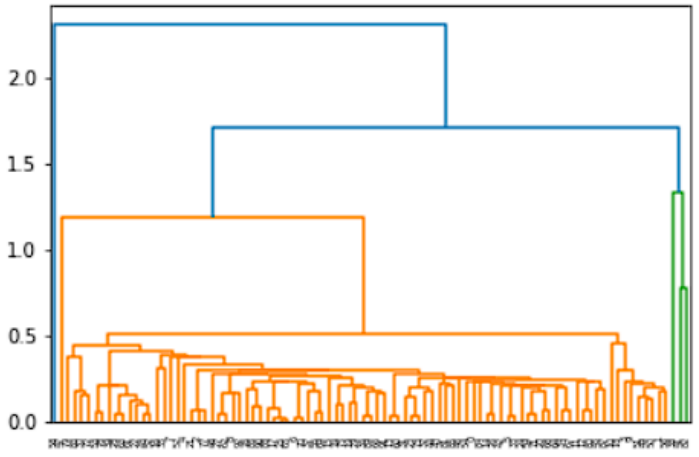
Hierarchical clustering (Customer Mall example)



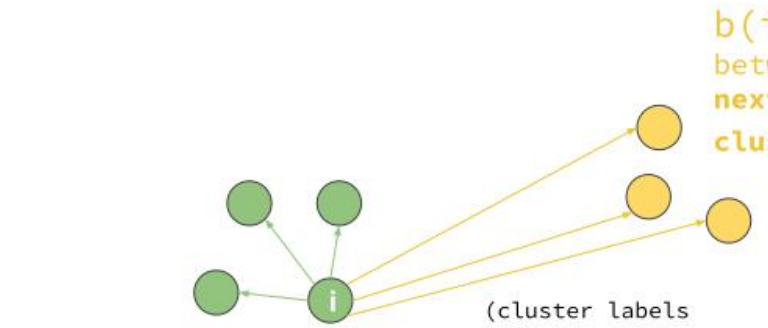
Agglomerative Clustering code

```
from sklearn.cluster import AgglomerativeClustering
agc = AgglomerativeClustering(n_clusters=7)
model = agc.fit(B1)
```

```
from scipy.spatial import distance_matrix
DM = distance_matrix(B1,B1)
from scipy.cluster.hierarchy import linkage
linkages = linkage(DM)
from scipy.cluster.hierarchy import dendrogram
Q = dendrogram(linkages)
```



Silhoutte score calculation



b(i): distance between i and its **next nearest cluster centroid**

(cluster labels assigned from ground-truth classes)

a(i): average distance between i and all of the other points in **its own cluster**

For a single point, i

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

$$-1 \leq s(i) \leq 1$$

Sprawling, overlapped clusters

Tight, well-separated clusters

Thank you

UTKARSH GAIKWAD