Unsupervised Learning: Clustering methods

Parcours Progis

Etudes, Medias, communication, Marketing

Bahareh Afshinpour

31.03.2025

Université Grenoble Alpes

Sciences Po Grenoble UGA

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Suggested Reading

- Introduction to Machine Learning with Python, A GUIDE FOR DATA SCIENTISTS, Andreas C. Müller & Sarah Guido ,oreilly

- https://medium.com/@esha.jagatia/k-means-clustering-in-social-media-management-a-statistical-approach-8284427b5ed1

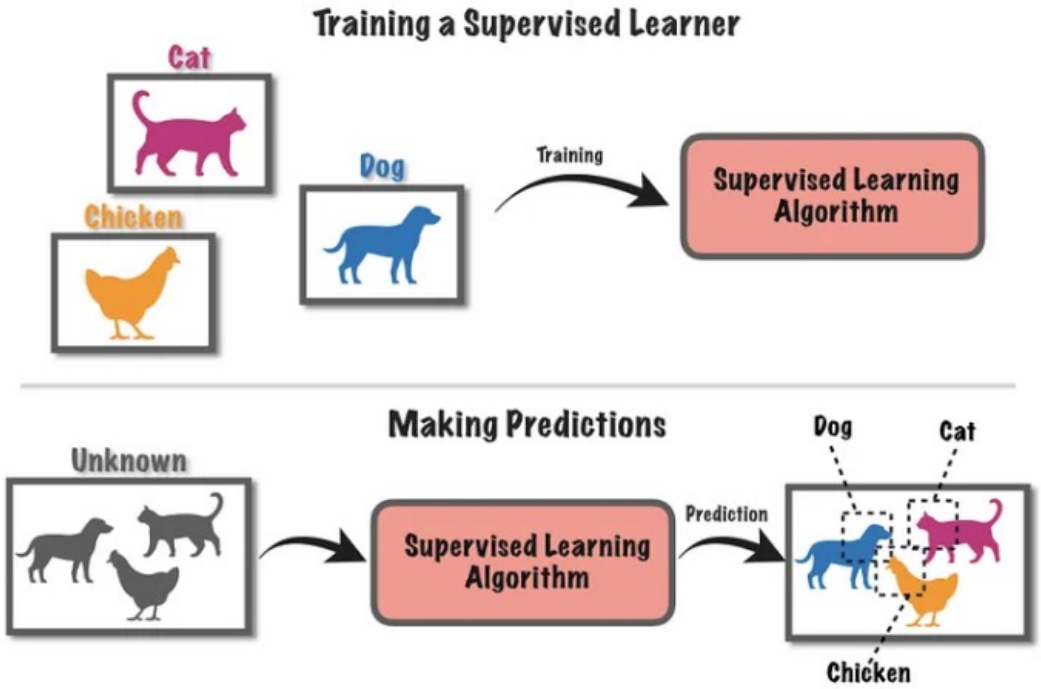
PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Supervised vs unsupervised

- In **supervised learning**, the aim is to learn a mapping from the input to an output whose correct values are provided by a supervisor.
- In **unsupervised learning**, there is no such supervisor and we only have input data.The aim is to find the regularities in the input.

Unsupervised Learning involves analysing **unlabeled data** to uncover hidden patterns or structures whithin data.

Supervised learning



https://kasunprageethdissanayake.medium.com/artificial-intelligence-2-supervised-learning-unsupervised-learning-and-reinforcement-learning-7bf00c732e99

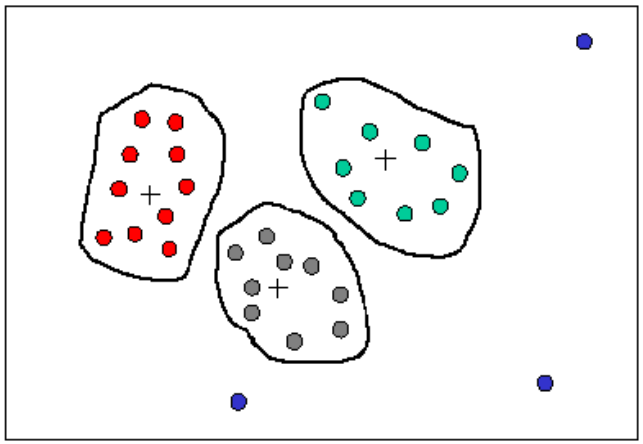
- Unsupervised learning subsumes all kinds of machine learning where there is no known output, no teacher to instruct the learning algorithm.
- In unsupervised learning, the learning algorithm is just shown the input data and asked to extract knowledge from this data.

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

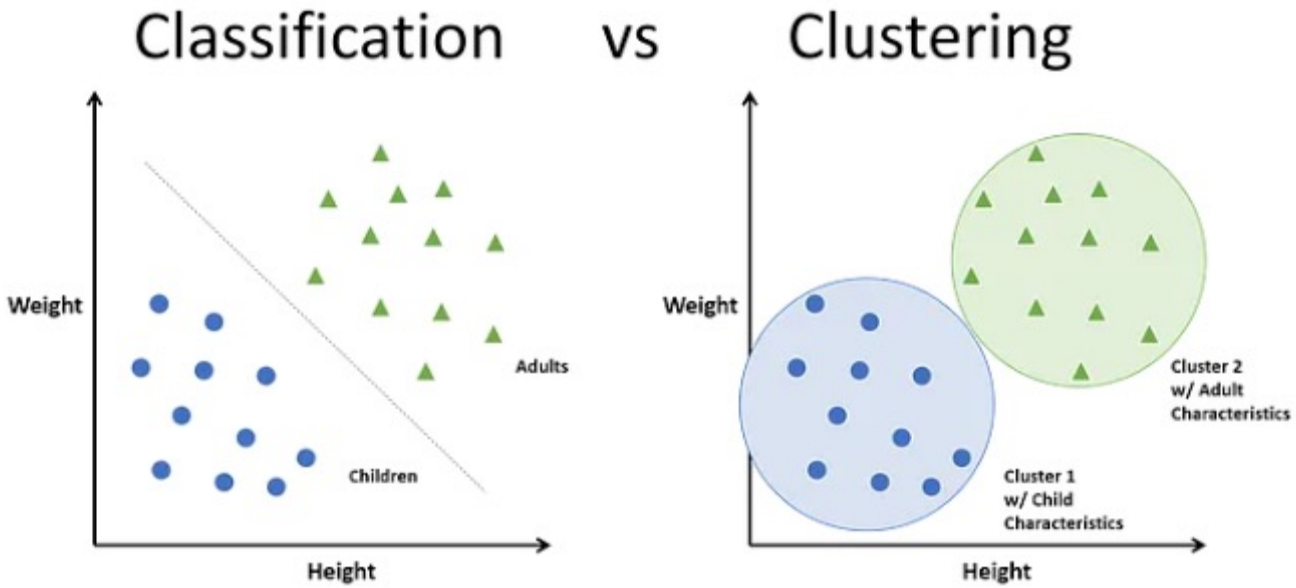
Clustering

Clustering is a process of partitioning a set of data (or objects) in a set of meaningful sub-classes, called clusters

- Cluster:
  - a collection of data objects that are “similar” to one another and thus can be treated collectively as one group
  - but as a collection, they are sufficiently different from other groups

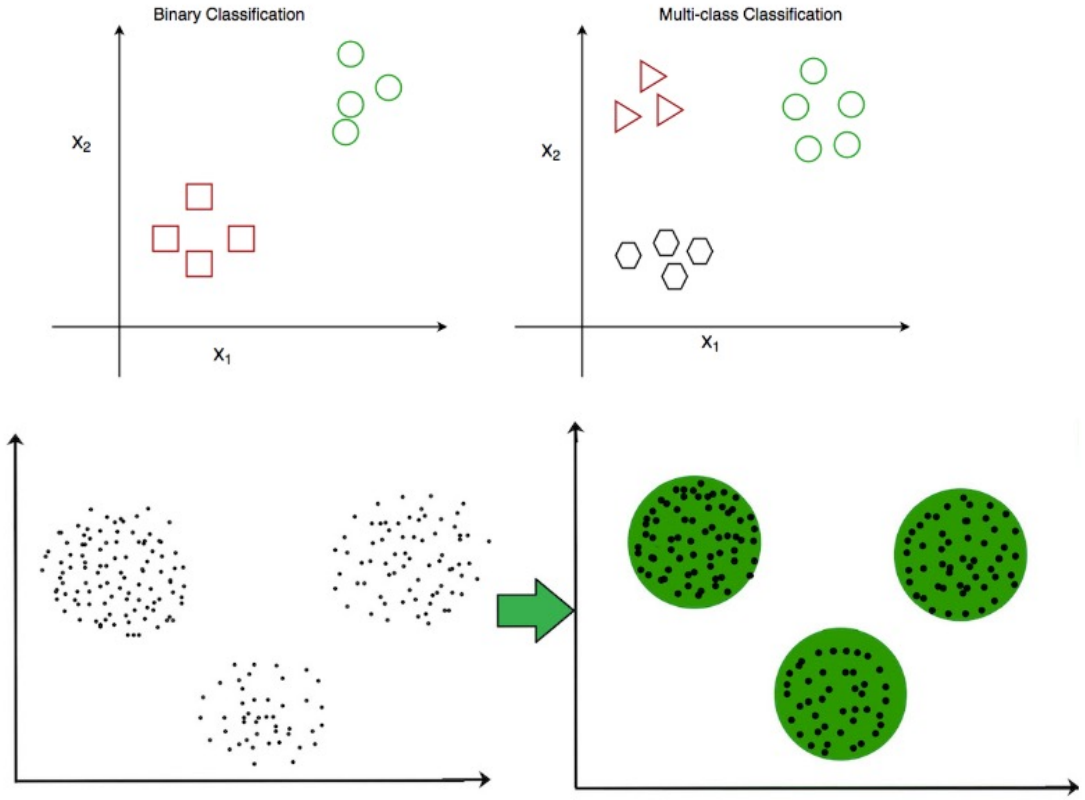


Classification vs Clustering

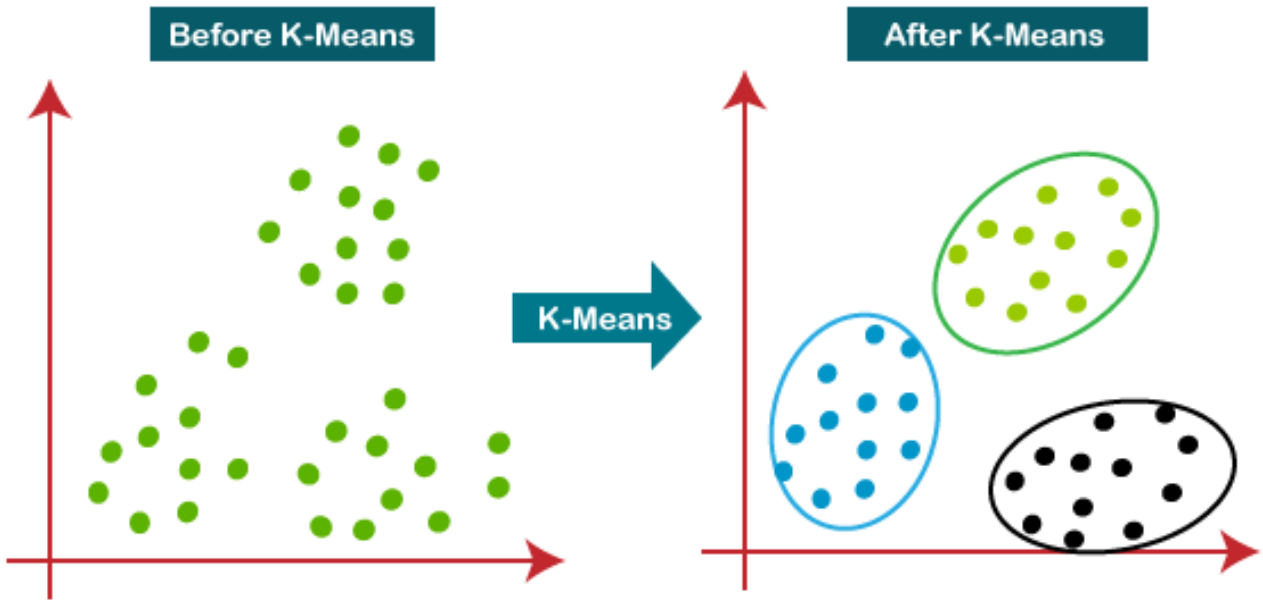


https://kevin-c-lee26.medium.com/machine-learning-101-classification-vs-clustering-e11b12c71243

Classification vs Clustering



https://www.geeksforgeeks.org/ml-classification-vs-clustering/



https://www.kaggle.com/code/mohammadbolandraftar/cluster-analysis-using-python-k-means

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Classification vs Clustering

Parameter	Classification	Clustering
Type	used for supervised learning	used for unsupervised learning
Basic	process of classifying the input instances based on their corresponding **class labels**	grouping the instances based on their **similarity** without the help of class labels
Need	it has labels so there is need of **training and testing** dataset for verifying the model created	there is no need of training and testing dataset
Example Algorithms	Logistic regression, K-NN, MLP, Decision Tree, Support vector machines, etc.	k-means clustering, agglomerative clustering, etc.

https://www.geeksforgeeks.org/ml-classification-vs-clustering/

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Application of clustering

Where to use clustering?

- ✓Document clustering
- ✓Emotion detection
- ✓Customer segmentation (Marketing)
- ✓Image Processing and object detection
- ✓Anomaly detection
- ✓Genomics and bioinformatics
- ✓Social network analysis and community detection

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Some common tasks

- We will look into two kinds of unsupervised learning in this course:
  - transformations of the dataset and clustering.

1. Clustering: grouping datapoints into clusters based on similarity

1. Unsupervised transformations of a dataset are algorithms that create a new representation of the data which might be easier for humans or other machine learning algorithms to understand.
   - dimensionality reduction

Clustering Methodologies

- Two general methodologies
  - Partitioning Based Algorithms
  - Hierarchical Algorithms
- Partitioning Based
  - divide a set of N items into K clusters (top-down)
- Hierarchical
  - agglomerative: pairs of items or clusters are successively linked to produce larger clusters

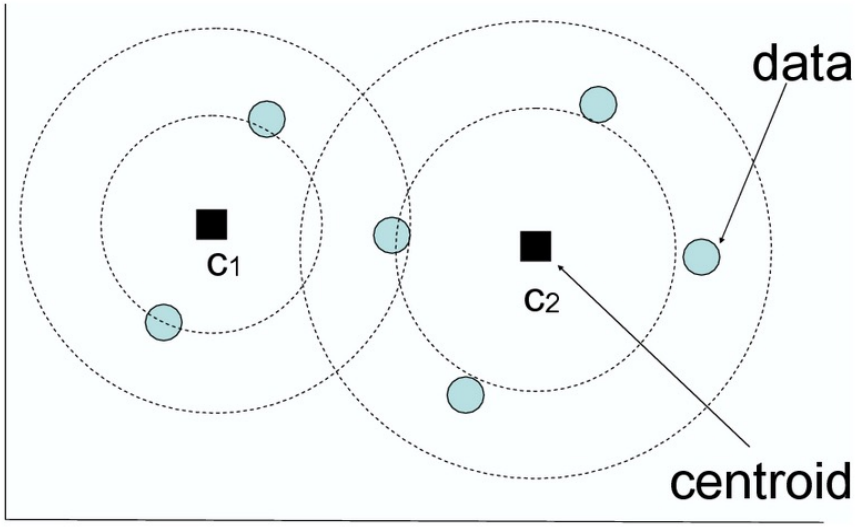
PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

K-means clustering

- The most widely used clustering algorithm.
- It partitions data into K distinct group based on feature similarity.

K-Means Clustering

- Separate the objects (data points) into K clusters.
- Cluster center (centroid) = the average of all the data points in the cluster.
- Assigns each data point to the cluster whose centroid is nearest (using distance function.)



PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

K-Means Algorithm

1. Assign K random points in the feature space as initial cluster centroids.

2. Based on distance from K centroids, assign nearest points to clusters.

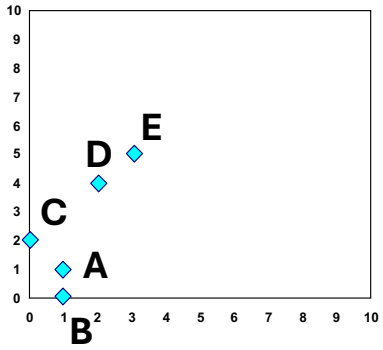
$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

1. Recalculate the positions of the K centroids.

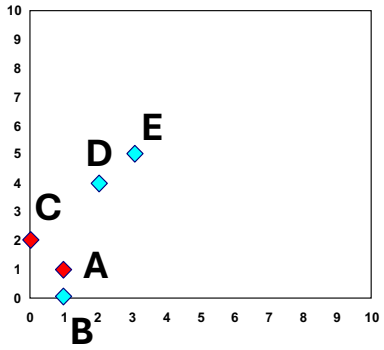
2. Repeat Steps 2 & 3 until the group centroids no longer move.

Example

i	x1	X2
A	1	1
B	1	0
C	0	2
D	2	4
E	3	5

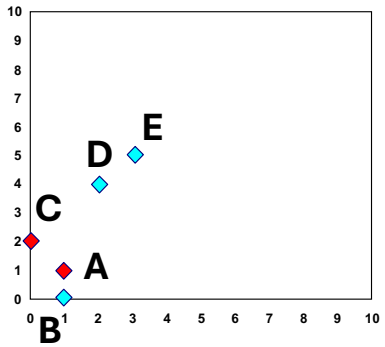


We can pick A and C as the two center to start.



Example

i	x1	X2
A	1	1
B	1	0
C	0	2
D	2	4
E	3	5



Let's figure out how far away (distance) all of these points are from A and C:

Center1=A=(1,1)

Center2=C=(0,2)

A- Center1=sqrt((1-1)^2+(1-1)^2)=0

A- Center2=sqrt((1-0)^2+(1-2)^2)=1.4

B- Center1=sqrt((1-1)^2+(1-0)^2)=0

B- Center2=sqrt((1-0)^2+(0-2)^2)=2.2

B is near Ceter1(A)

C- Center1=sqrt((0-1)^2+(2-0)^2)=2.2

C- Center2=sqrt((0-0)^2+(2-2)^2)=2

D- Center1=sqrt((2-1)^2+(4-0)^2)=sqrt(17)=4.1

D- Center2=sqrt((2-0)^2+(4-2)^2)=sqrt(8)=2.8

D is near Ceter2(C)

E- Center1=sqrt((3-1)^2+(5-0)^2)=sqrt(29)=5.3

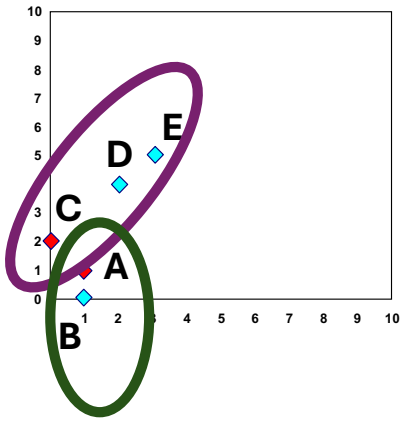
E is near Ceter2(C)

E- Center2=sqrt((3-0)^2+(5-2)^2)=sqrt(18)=4.2

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Example

i	x1	X2	cluster
A	1	1	1
B	1	0	1
C	0	2	2
D	2	4	2
E	3	5	2



We need to find the average of each group, which is the average of all those memebrs in that group.

Center1= avg A,B=((1,1)+(1,0))/2= (1,0.5)

Center2=AVG C,D,E=((0,2),(2,4),(3,5))/3= (1.7, 3.7)

A- Center1=sqrt((1-1)^2+(1-0.5)^2)=0.5

A- Center2=sqrt((1-1.7)^2+(1-3.7)^2)=2.8

A is near Ceter1

B- Center1=sqrt((1-1)^2+(1-0.5)^2)=0.5

B- Center2=sqrt((1-1.7)^2+(0-3.7)^2)=3.7

B is near Ceter1

C- Center1=sqrt((0-1)^2+(2-0.5)^2)=1.8

C- Center2=sqrt((0-1.7)^2+(2-3.7)^2)=2.4

C is near Ceter1

D- Center1=sqrt((2-1)^2+(4-0.5)^2)=3.6

D- Center2=sqrt((2-1.7)^2+(4-3.7)^2)=0.42

D is near Ceter2

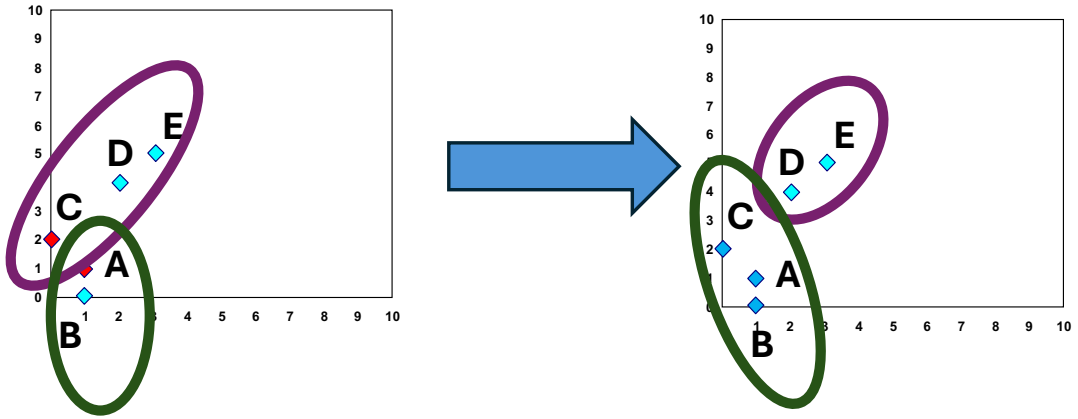
E- Center1=sqrt((3-1)^2+(5-0.5)^2)=4.9

E- Center2=sqrt((3-1.7)^2+(5-3.7)^2)=1.8

E is near Ceter2

Example

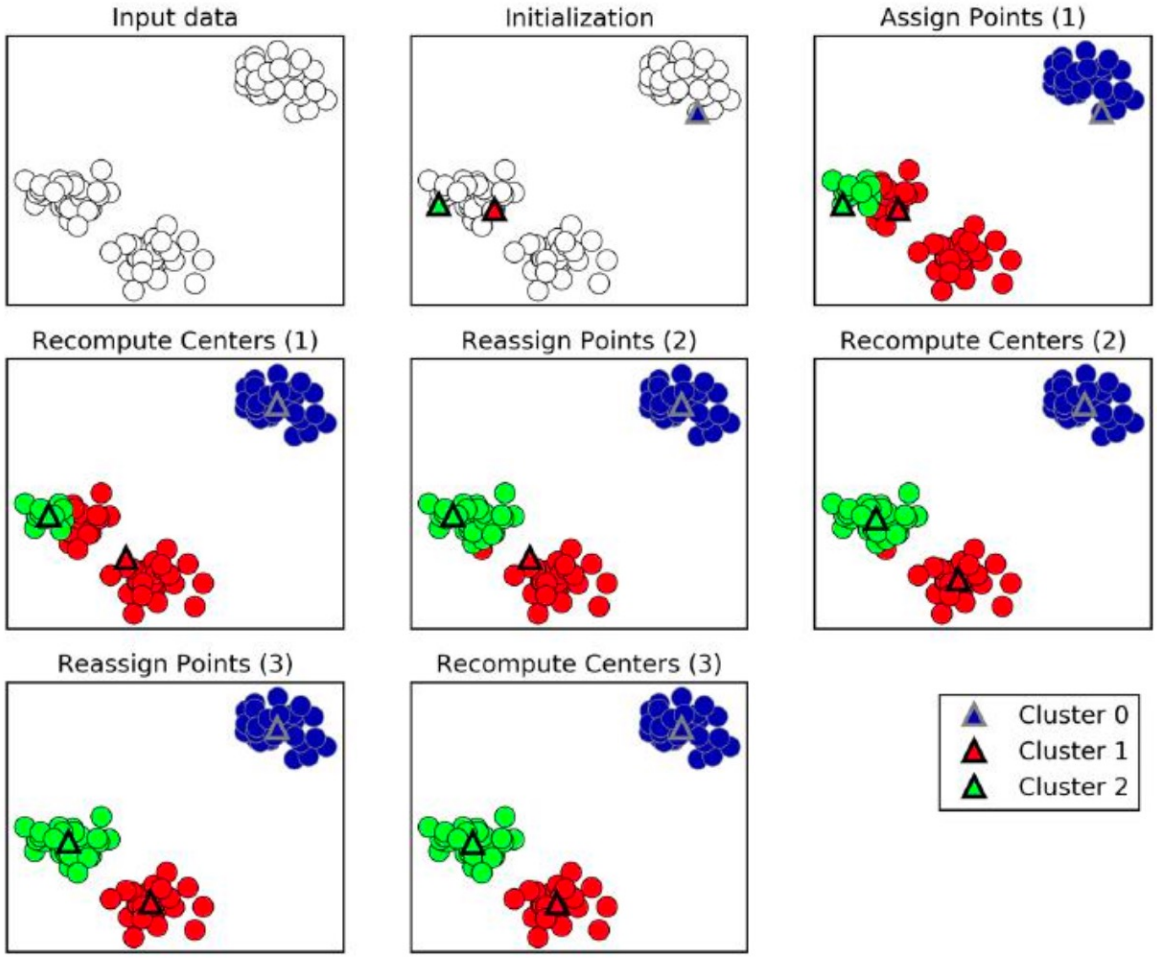
i	x1	X2	cluster
A	1	1	1
B	1	0	1
C	0	2	1
D	2	4	2
E	3	5	2



PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Example

- We keep going until we get the same result as our previous step. Since it doesn't seem to have changed, we can stop and report the clusters.


Input data
Initialization
Assign Points (1)
Recompute Centers (1)
Reassign Points (2)
Recompute Centers (2)
Reassign Points (3)
Recompute Centers (3)
Cluster 0
Cluster 1
Cluster 2

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Questions

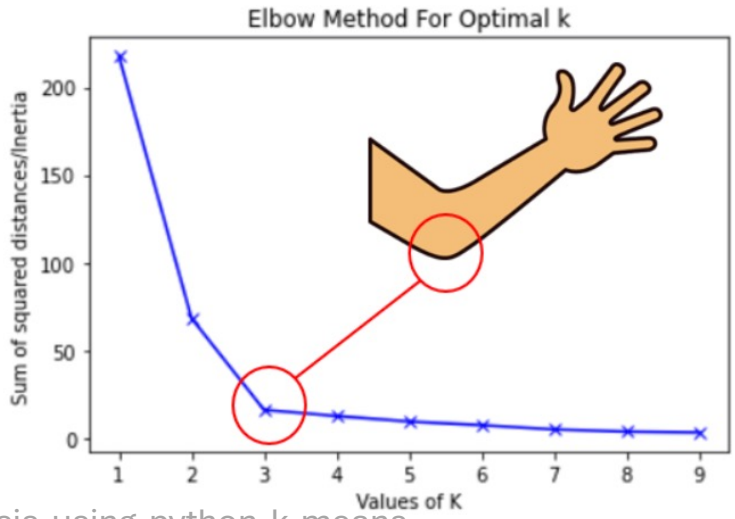
- How to create ‘good’ clusters?
- How many clusters do we need?

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

How many clusters – Elbow method

- Find the optimal number of clusters K by minimizing the within-cluster sum of squares
- Elbow method:
  - The point where the rate of decrese sharply slows down (resembles an « elbow») is considered the optimal K.
  - The elbow method calculates

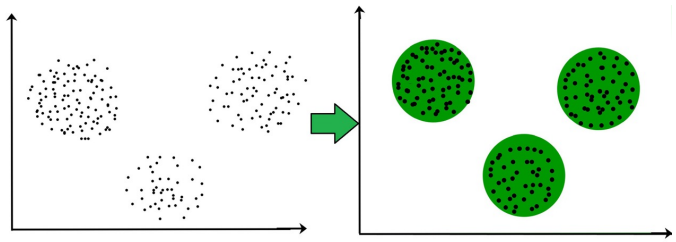
the sum of the square of the points and calculates the average distance.



https://www.kaggle.com/code/mohammadbolandraftar/cluster-analysis-using-python-k-means

What Is Good Clustering?

- A good clustering method will produce high quality clusters with
  - Low intra-class similarity
  - High inter-class similarity
- The quality of a clustering result depends on both the similarity measure used by the method and its implementation.
- The quality of a clustering method is also measured by its ability to discover some or all of the hidden patterns.



https://www.geeksforgeeks.org/clustering-in-machine-learning/

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Objective of k means Clustering

- The main objective of k-means clustering is to partition your data into a specific number (k) of groups
- **Grouping similar data points**
- **Minimizing within-cluster distance**
- **Maximizing between-cluster distance**

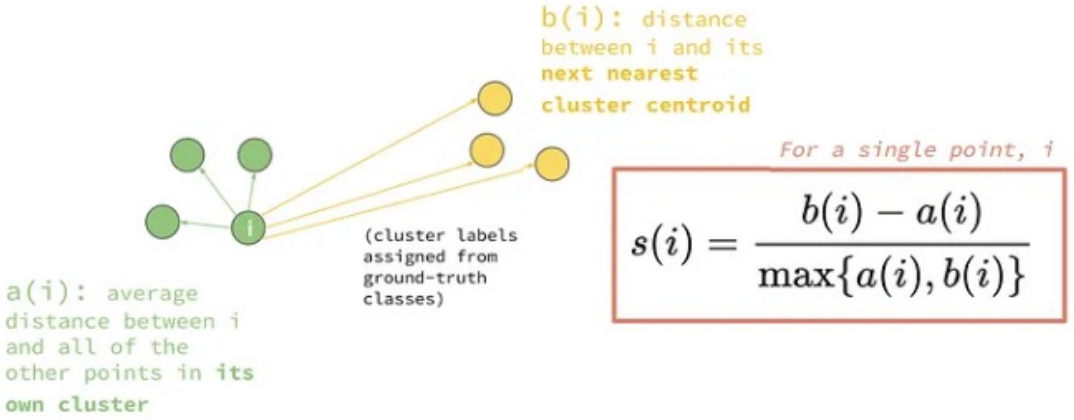
PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Evaluation - Silhouette Score

- It provides a quantitative way to measure how well data points are grouped and separated into clusters.
- **+1**: Perfect clustering — the point is far from other clusters and well-matched to its own cluster.
- **0**: Borderline clustering — the point is equidistant between two clusters.
- **-1**: Poor clustering — the point is closer to another cluster than its own.

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Silhouette Score



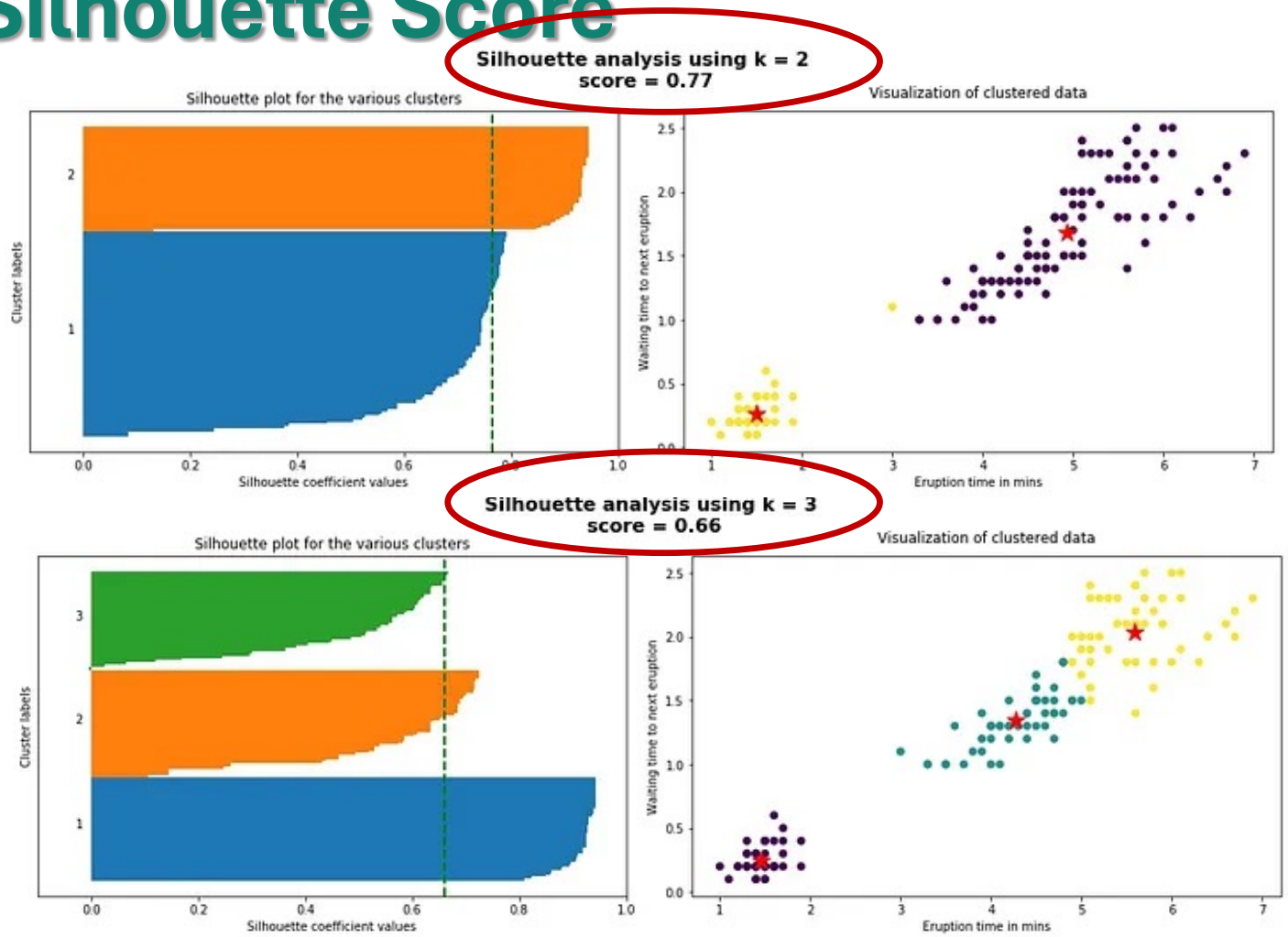
It considers two aspects:

1. Cohesion (a(i)): How close a data point is to other points in its own cluster.

2. Separation (b(i)): How far a data point is from the points in the nearest neighboring cluster.

https://farshadabdulazeez.medium.com/understanding-silhouette-score-in-clustering-8aedc06ce9c4

Silhouette Score



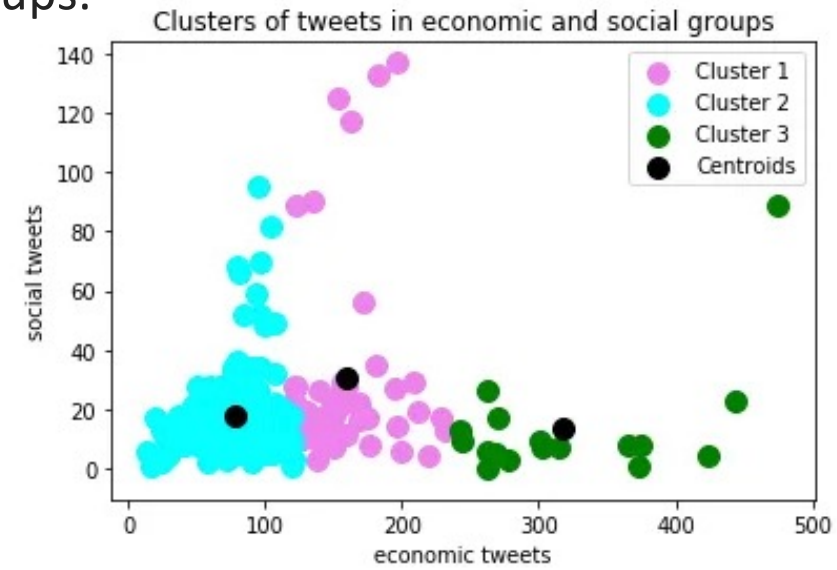
we can clearly see that how plot and score are different according to n_cluster(k) .
So, we can easily choose **high score** and **number of k** via silhouette analysis technique

https://medium.com/@cmukesh8688/silhouette-analysis-in-k-means-clustering-cefa9a7ad111

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Example of k-Means-Tweets Clustering

- 4 major categories: *Economic, Social, Cultural and Health* then performing KMeans cluster analysis on the groups.



https://medium.com/swlh/tweets-classification-and-clustering-in-python-b107be1ba7c7

The plot above indicates most of the users share more economy-centred tweets compared to social tweets. There's a few who try to maintain a balance between the categories.

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Applications of K-Means in Social Media Management

- **Audience Segmentation**

  Social media platforms cater to diverse audiences with different interests and behaviors. K-means clustering helps group users based on: Age, gender, location, Likes, shares, comments interests and Common topics they interact with.

- **Hashtag and Content Strategy Optimization**

  Marketers use K-means to analyze hashtags and content performance by clustering posts based on: Engagement metrics (likes, shares, comments),Content themes (tech, fitness, fashion, etc.) and Posting time.

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Applications of K-Means in Social Media Management

- **Influencer Identification**

  K-means helps businesses identify and categorize influencers based on: Engagement rates , Follower demographics. This ensures brands collaborate with the right influencers for maximum impact.

- **Sentiment Analysis and Customer Insights**
  - K-means clustering helps analyze user comments, reviews, and discussions to track sentiment trends. By grouping text data into positive, neutral, and negative clusters, brands can Improve customer service.

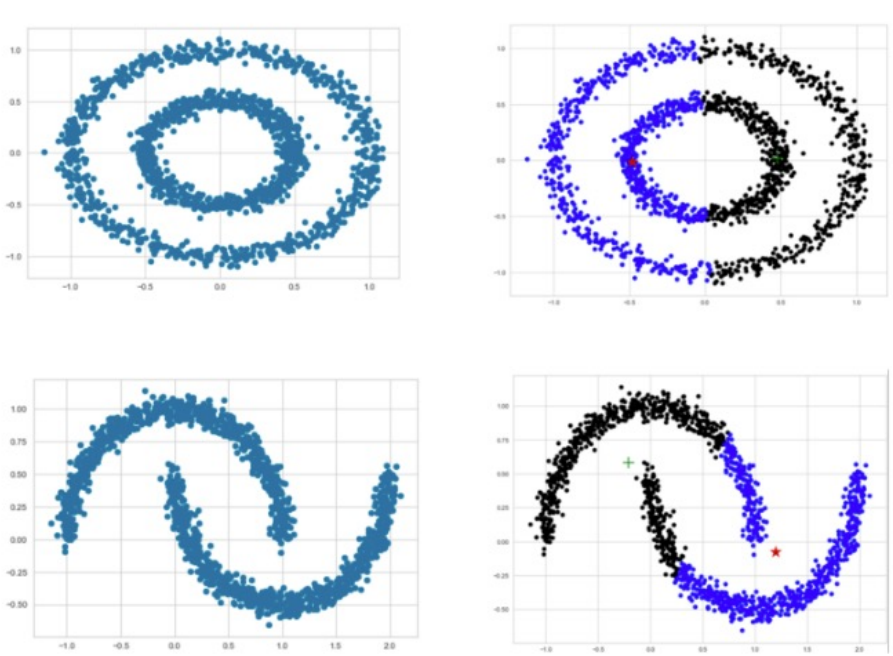
PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

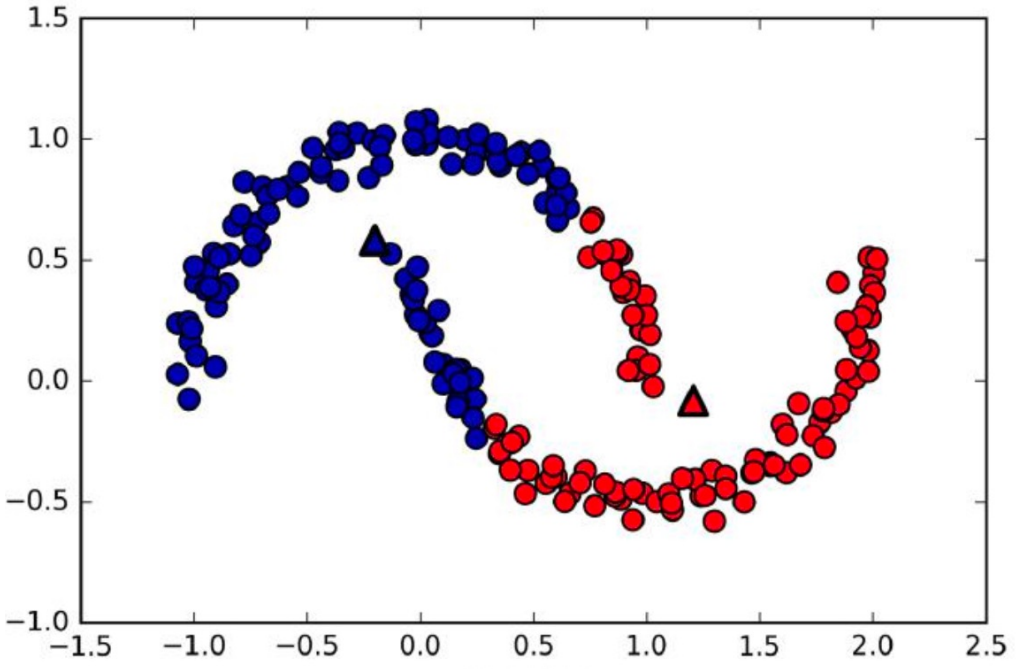
Challenges - Data distribution

- When using **K-means clustering** on complex distributions, there are a key challenge:
- **Assumption of spherical clusters**: K-means assumes that clusters are spherical (or roughly circular). This means it struggles when the data has clusters with different shapes, densities, or sizes (e.g., elongated, concentric, or overlapping clusters).
- If your data consists of non-convex clusters (like crescent shapes or clusters with holes), K-means will struggle to identify them correctly.

Challenges - Data distribution

- the two_moons data





https://towardsdatascience.com/explain-ml-in-a-simple-way-k-means-clustering-e925d019743b/

Introduction to Machine Learning with Python, chapter 3

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

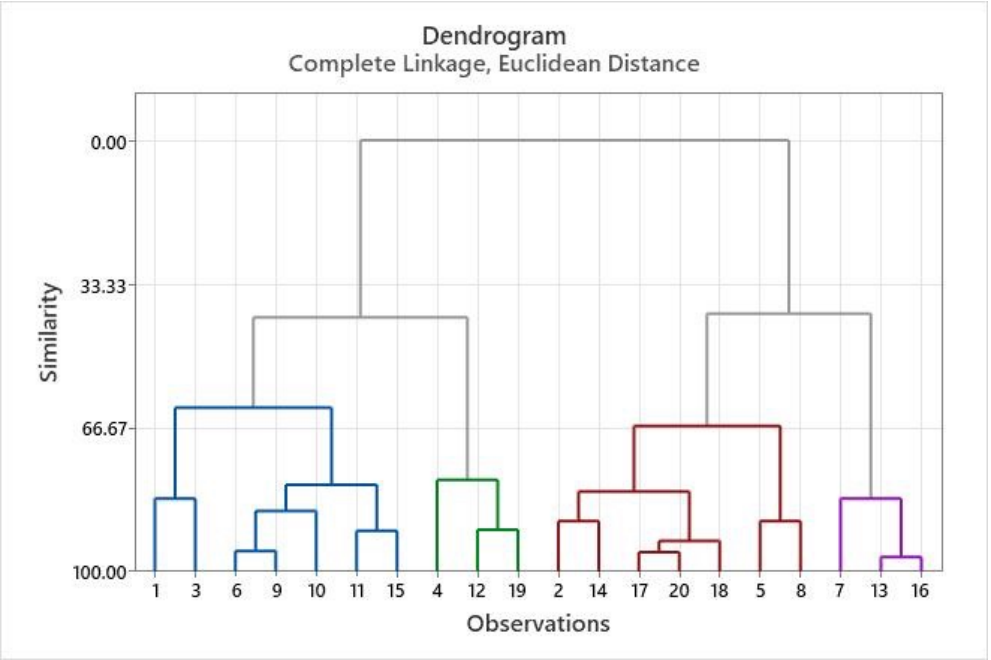
Challenges - Data distribution

Agglomerative Clustering: Can be more flexible than K-means and handle complex shapes, though still limited with respect to density differences.

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Hierarchical clustering

- It is a Clustering method that create a **hierarchical tree** of objects.
- The tree represents the relationships between the objects.
- It is Bottom-up approach:
  - Sice each observation consider as a cl
  - Then they combine together based on their similarities to form clusters
- The result of a hierarchical clustering is called a dendogram



https://support.minitab.com/en-us/minitab/help-and-how-to/statistical-modeling/multivariate/how-to/cluster-observations/interpret-the-results/all-statistics-and-graphs/dendrogram/

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

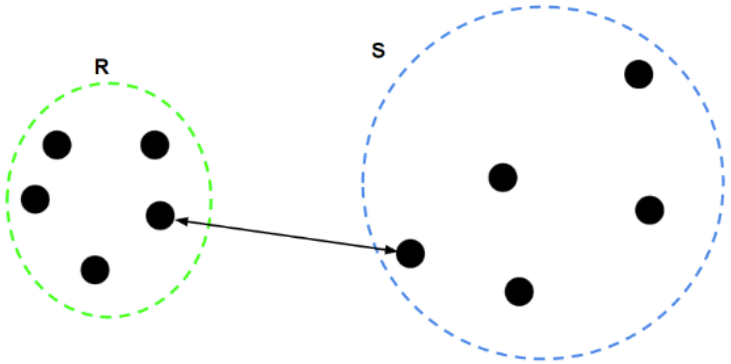
Steps of hierarchical clustering

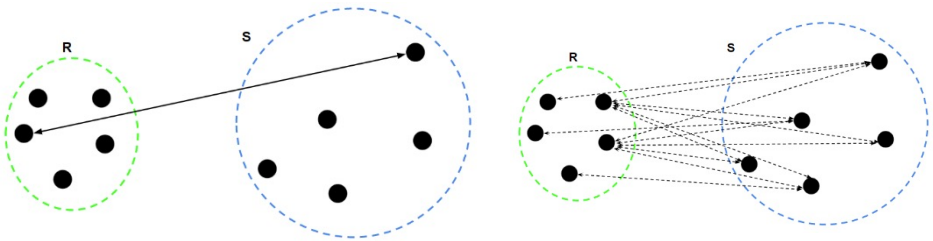
- Start by assigning each item to a cluster
- Let the distances (similarities) between the clusters the same as the distances (similarities) between the items they contain.
- Find the closest (most similar) pair of clusters and merge them into a single cluster, so that now you have one cluster less.
- Compute distances (similarities) between the new cluster and each of the old clusters.
- Repeat steps 2 and 3 until all items are clustered into K number of clusters

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Measuring similarity between clusters

- There are some method to calculate distances between clusters.
- The most commonly use method is **single linkage** method.
- If you have two clusters, the distance between these two, is the shortest possible distance
- There are other method like :
  - **Average Linkage**
  - **And complete linkage**





Example of Hierarchical clustering

Distances in kilometers between some italian cities.

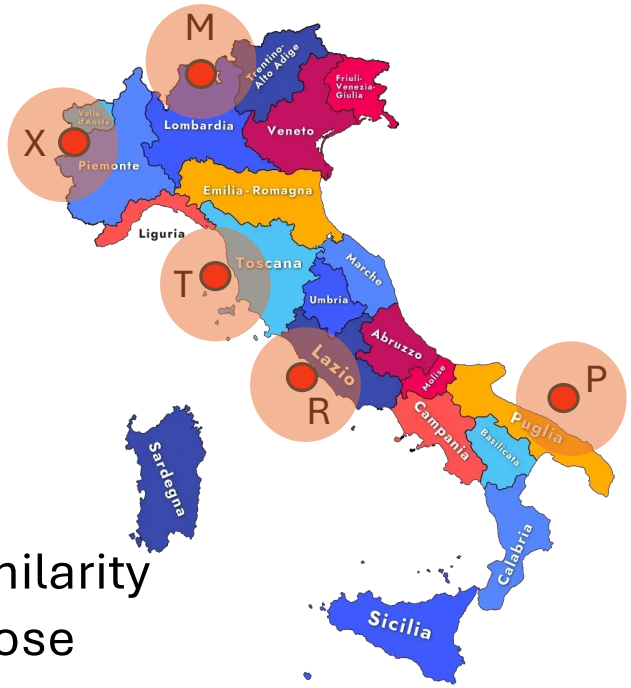
	P	T	M	R	X
P	0	662	877	412	996
T	662	0	295	268	400
M	877	295	0	564	138
R	412	268	564	0	669
X	996	400	138	669	0

The goal here is to find cities that are closest.

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Example of Hierarchical clustering

	P	T	M	R	X
P	0	662	877	412	996
T	662	0	295	268	400
M	877	295	0	564	138
R	412	268	564	0	669
X	996	400	138	669	0



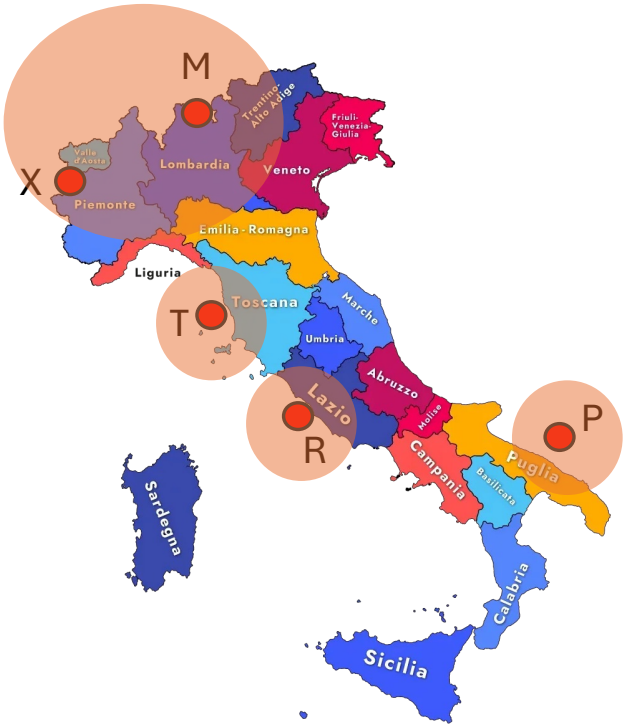
-Here, first we assume each city is the cluster
-Then we start to put cities together based on thier similarity
-distance between x and M is 139 and they are very close
-So we have new cluster (M,X)
- We should update deistance matrix

Example of Hierarchical clustering

	P	T	M	R	X
P	0	662	877	412	996
T	662	0	295	268	400
M	877	295	0	564	138
R	412	268	564	0	669
X	996	400	138	669	0

	P	T	M/X	R
P	0	662		412
T	662	0		268
M/X			0	
R	412	268		0

I eliminated one row and one column

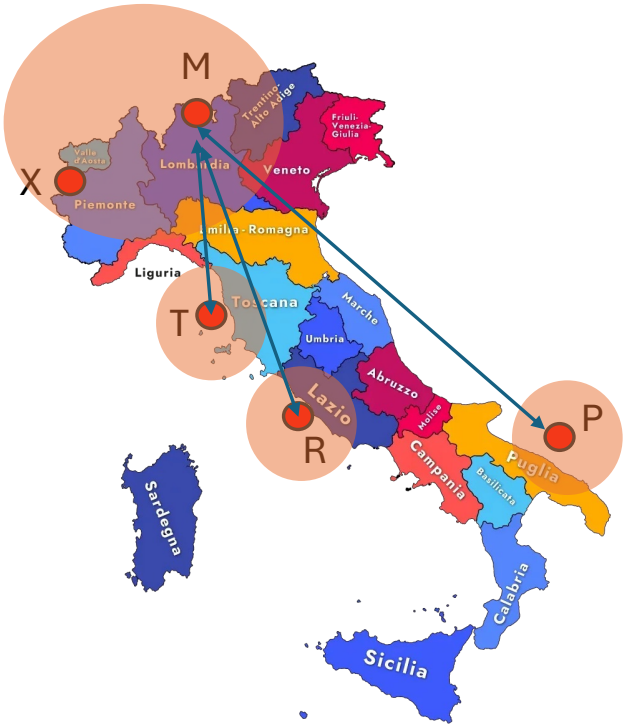


- We combine M and X and have a new cluster
- We need to calculate the distance from other cities to this new cluster.

Example of Hierarchical clustering

	P	T	M	R	X
P	0	662	877	412	996
T	662	0	295	268	400
M	877	295	0	564	138
R	412	268	564	0	669
X	996	400	138	669	0

	P	T	M/X	R
P	0	662	**877**	412
T	662	0	**295**	268
M/X	**877**	**295**	0	564
R	412	268	564	0

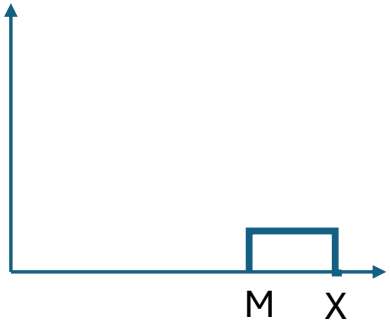


- Single linkage

Example of Hierarchical clustering

	P	T	M/X	R
P	0	662	**877**	412
T	662	0	**295**	268
M/X	**877**	**295**	0	564
R	412	268	564	0

- We have to continue this process of combining
- And in each step we can create initial dendogram



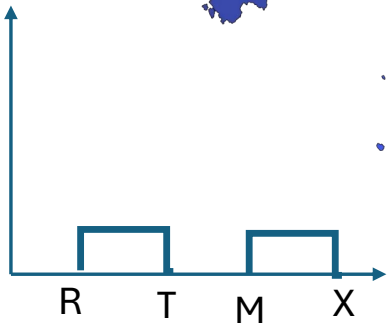
PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Example of Hierarchical clustering

	P	T	M/X	R
P	0	662	**877**	412
T	662	0	**295**	268
M/X	**877**	**295**	0	564
R	412	268	564	0

- What are the next two cities that are the closest together?

	P	T	M/X	R
P	0	662	**877**	412
T	662	0	**295**	268
M/X	**877**	**295**	0	564
R	412	268	564	0

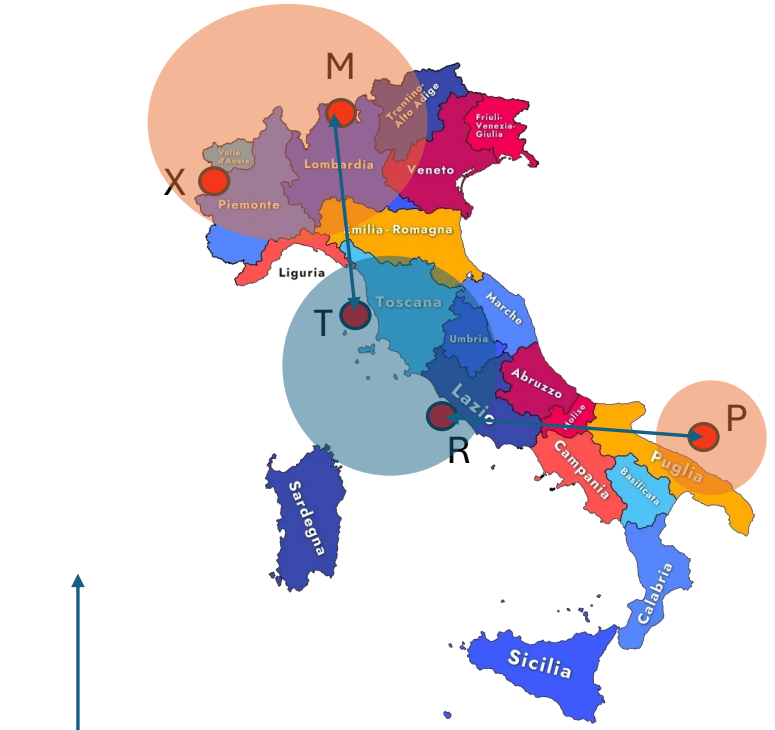


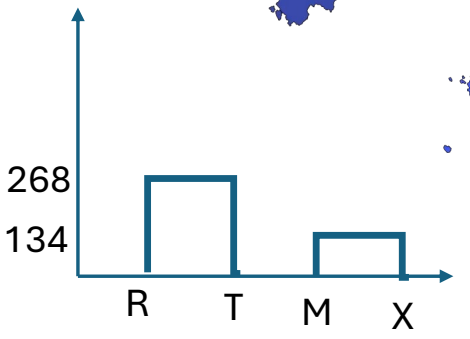
Example of Hierarchical clustering

	P	T	M/X	R
P	0	662	**877**	412
T	662	0	**295**	268
M/X	**877**	**295**	0	564
R	412	268	564	0

	P	R/T	M/X
P	0	412	**877**
M/X	**877**	**295**	0
R/T	412	0	295

- Know we have a new cluster (R/T)
- In each step we refer to previous matrix

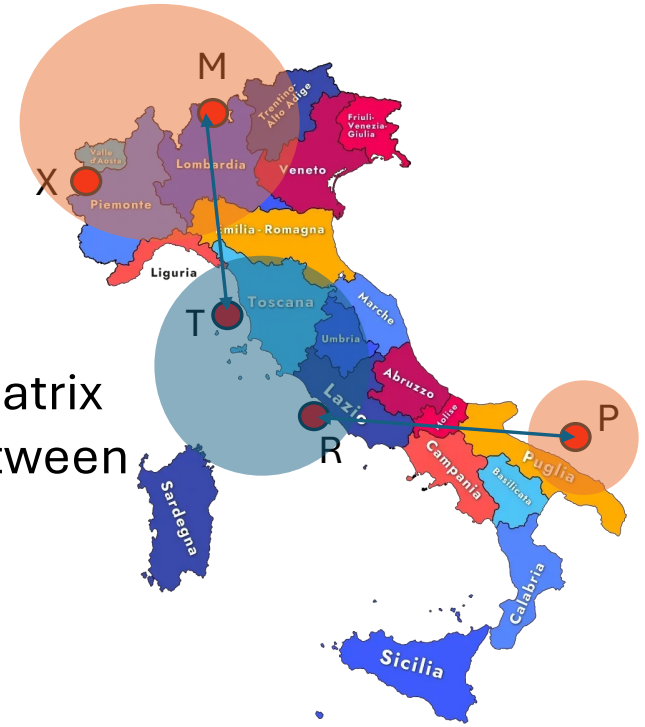




PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Example of Hierarchical clustering

	P	R/T	M/X
P	0	412	**877**
M/X	**877**	**295**	0
R/T	412	0	295

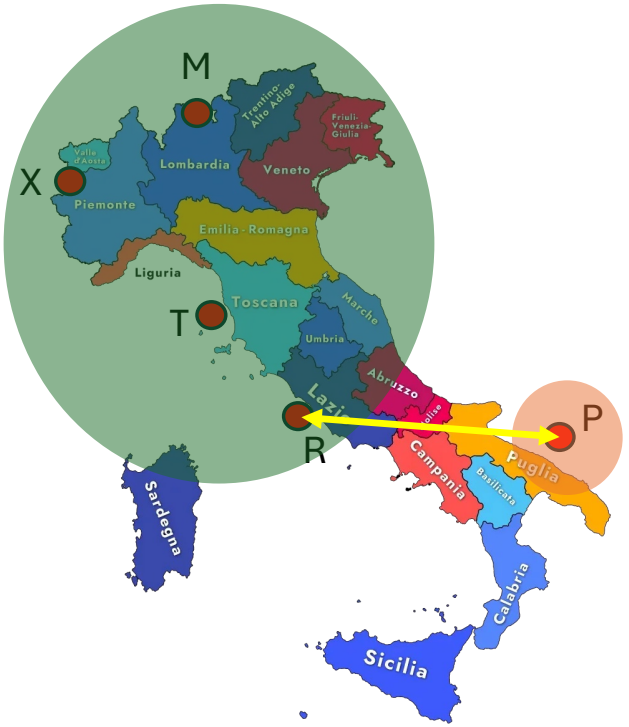


- We continue this process until we get to a 2 by 2 matrix
- To update matrix the minimum distance is now between M/X and R/T – 295
- So wecombine R/T and M/X cluster

Example of Hierarchical clustering

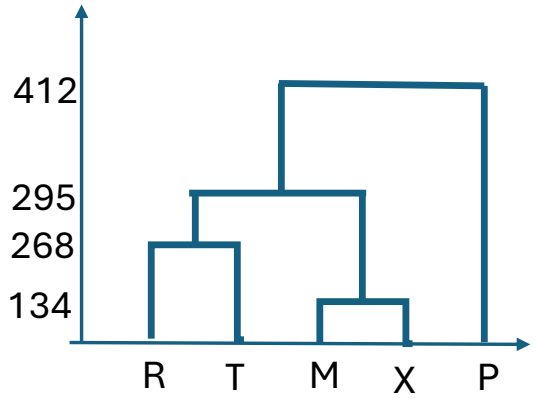
	P	R/T	M/X
P	0	412	**877**
M/X	**877**	**295**	0
R/T	412	0	295

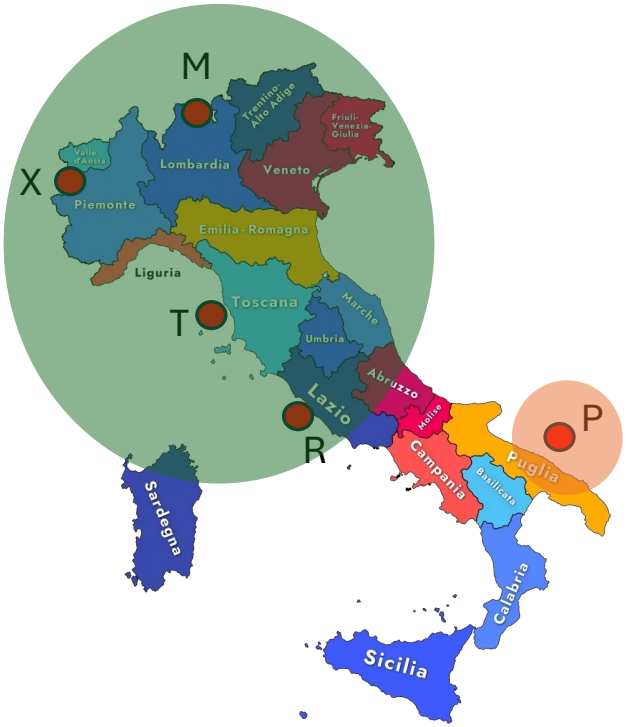
	P	R/T/M/X
P	0	412
R/T/M/X	412	**0**



- So we combine R/T and M/X cluster
- We update the matrix
- P is near to R (single linkage)

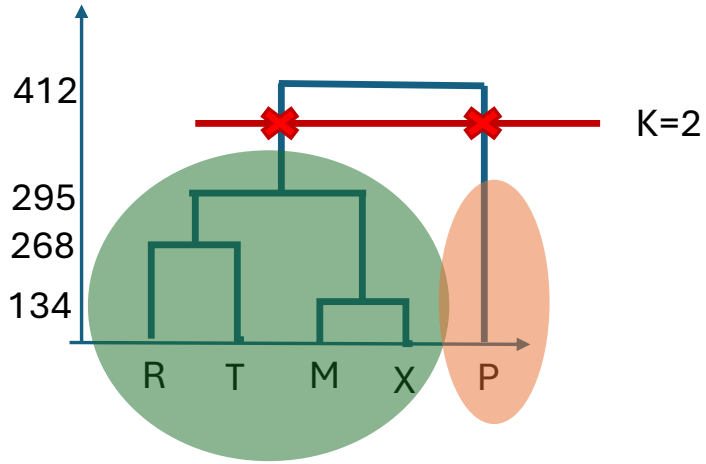
Example of Hierarchical clustering

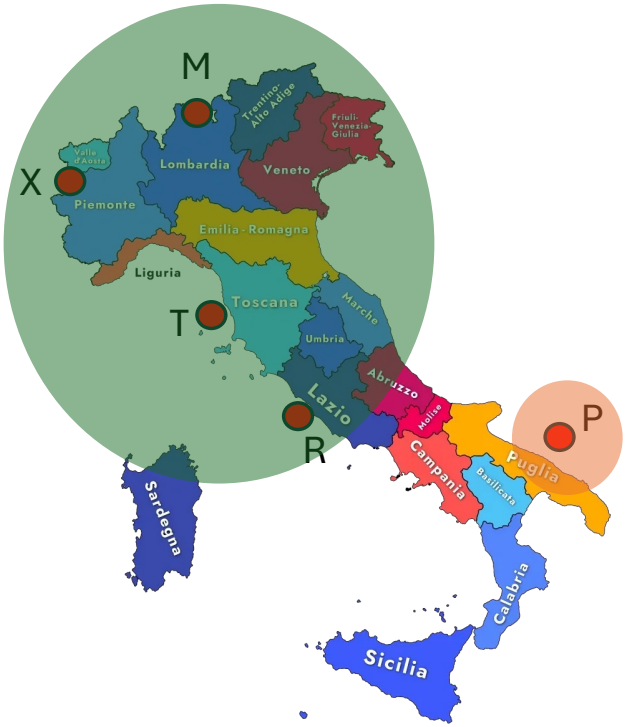




- We know that the length of this bar shows how far apart two clusters are from each other.
- To define the number of groups we have to create horizontal lines that hit the vertical lines.

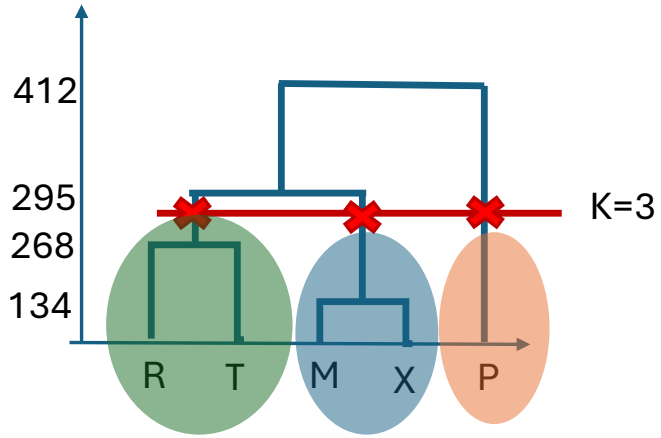
Example of Hierarchical clustering

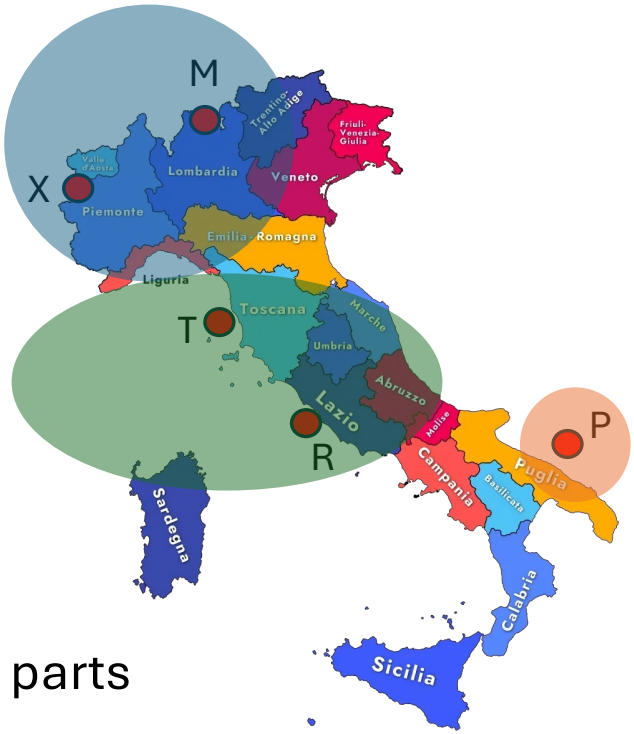




- We know that the length of this bar shows how far apart two clusters are from each other.
- To define the number of groups we have to create horizontal lines that hit the vertical lines.

Example of Hierarchical clustering



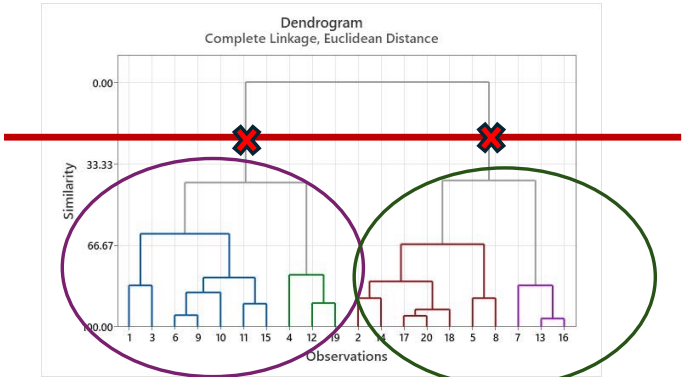


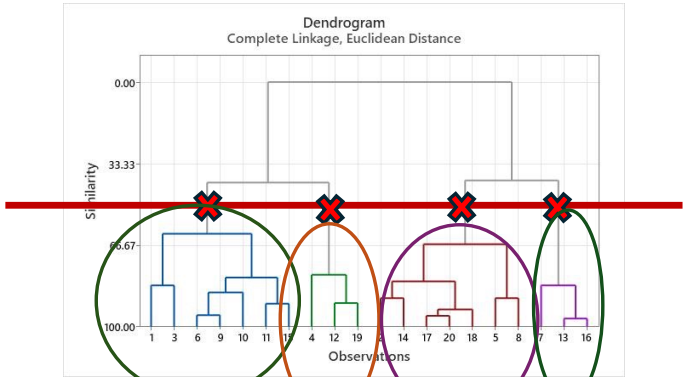
- If we want to have three group, we have to draw a horizontal line that hits these vertical line at three parts
- …
- **Good: We do not have predefined number of groups to start our clustering method.**

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Hierarchical clustering

- We know that the length of the bar in the dendogram shows how far apart two clusters are from each other.
  - The longer this length is, the more far apart two clusters are from each other.
- How can I use dendogram to perform the grouping of observations?
  - Create horizontal lines that hit the vertical lines
  - Each of the vertical lines lead us to a new cluster





PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Hierarchicla clustering versus k-means clustering

- ✓If there is a specific number of clusters in the dataset, but the group they belong to is unknown, choose K-means
- ✓With a **large number** of variables, **K-means** compute faster
- ✓It is easier to determine the number of clusters by hierarchical clustering's dendrogram

https://www.globaltechcouncil.org/data-science/k-means-clustering-vs-hierarchical-clustering/

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Hard vs soft clustering

- Hard clustering:
  - Eache data point belongs to exactly one cluser
  - More common and easier to use

- Soft clustering:
  - Each sample is assigned to different clusters with probabilities, rather than 0 and 1.
  - Data point belongs to each cluster with a probability

TP7

```
data='./Mall_Customers.csv'
df=pd.read_csv(data)
df.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

```
df.shape
```

```
(200, 5)
```

Checking the missing value

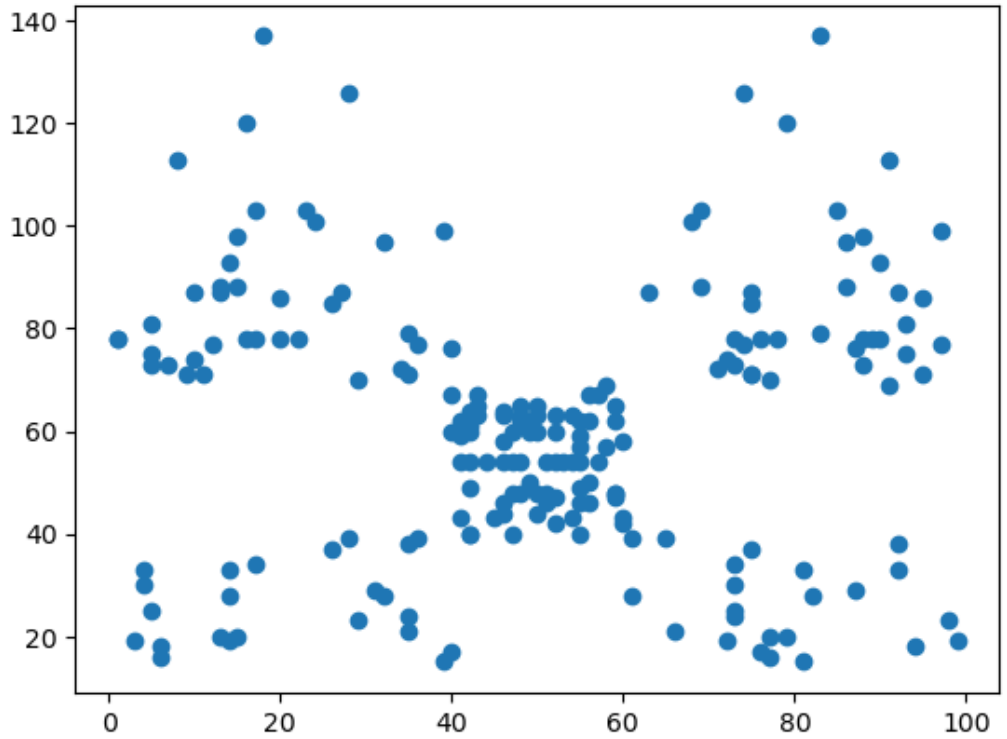
```
df.isna().sum()
```

```
CustomerID                0
Genre                     0
Age                       0
Annual Income (k$)        0
Spending Score (1-100)    0
cluster                   0
dtype: int64
```

Plot the data

```
plt.scatter(df["Spending Score (1-100)"],df["Annual Income (k$)"])
```

```
<matplotlib.collections.PathCollection at 0x157d9c110>
```



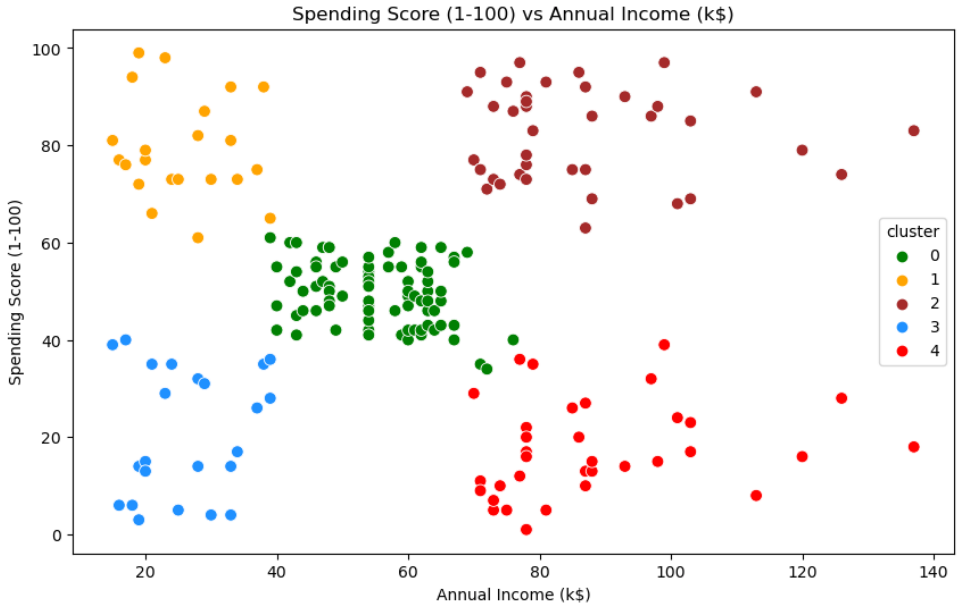
K-means clustering

```
km=KMeans(n_clusters=5)
y_predicted=km.fit_predict(df[["Spending Score (1-100)","Annual Income (k$)"]])
```

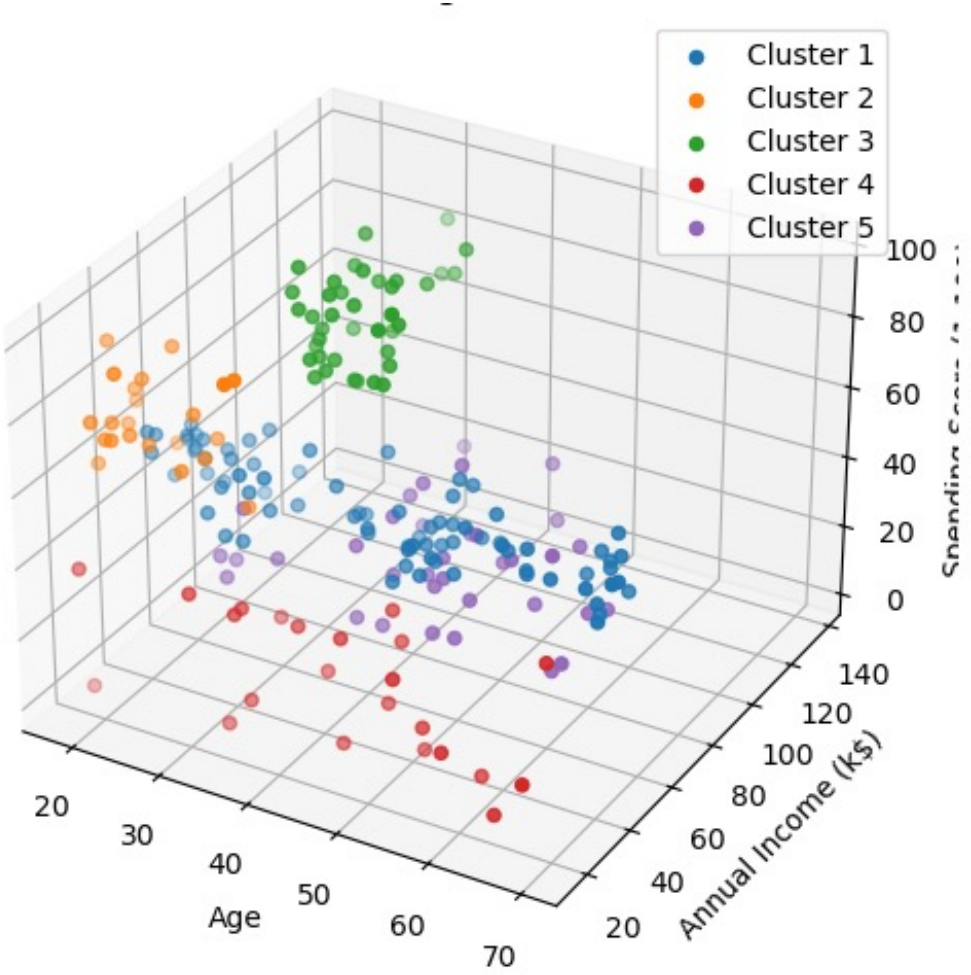
```
df["cluster"]=y_predicted
df.head()
```

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

```
import seaborn as sns
plt.figure(figsize=(10,6))
sns.scatterplot(x = 'Annual Income (k$)',y = 'Spending Score (1-100)',hue="cluster",
                palette=['green','orange','brown','dodgerblue','red'], legend='full',data = df  ,s = 60 )
plt.xlabel('Annual Income (k$)')
plt.ylabel('Spending Score (1-100)')
plt.title('Spending Score (1-100) vs Annual Income (k$)')
plt.show()
```



3D


Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Age
20
30
40
50
60
70
Annual Income (k$)
20
40
60
80
100
120
140
0
20
40
60
80
100

Silhouette score

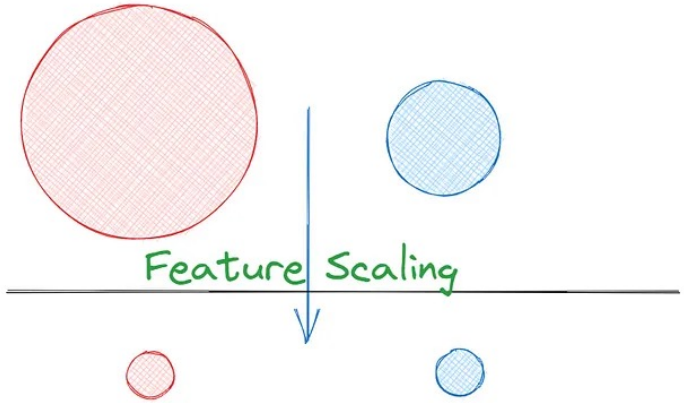
```
from sklearn.metrics import silhouette_score
scores = []
for k in range(2, 8):
    kmeans = KMeans(n_clusters=k, random_state=42)
    labels = kmeans.fit_predict(df[["Spending Score (1-100)","Annual Income (k$)"]])
    scores.append(silhouette_score(df[["Spending Score (1-100)","Annual Income (k$)"]], labels))
print(scores)
```

K=2, 0.2968969162503008,
K=3. 0.467613581587754 35,
K=4. 0.4931963109249047,
K=5. 0.553931997444648,
K=6. 0.53976103063432,
K=7. 0.5264283703685728]

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Feature Scaling

- Imagine you’re comparing apples and oranges; it's nonsensical.
- Features measured in vastly different units (like income and age) can skew machine learning algorithms if not addressed.
- Feature scaling addresses this by transforming data into a standard scale, enabling fair comparison between different features.



https://medium.com/@yashbaravaliya206/feature-scaling-and-normalization-ca484a16882a

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Feature Scaling

- Feature scaling is a fundamental preprocessing step in machine learning aimed at ensuring that numerical features have a similar scale.
- Apply feature scaling before feeding the data into the machine learning model (K-NN), **except** for algorithms that are scale-invariant, such as **decision trees**.

PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Common Techniques for Feature Scaling

- **Normalization :**
  - This method scales each feature so that all values are within the range of 0 and 1.
  - x is the value you want to normalize
  - min(X) is the minimum value in your data set
  - max(X) is the maximum value in your data set

$$(x - \min(X)) / (\max(X) - \min(X))$$

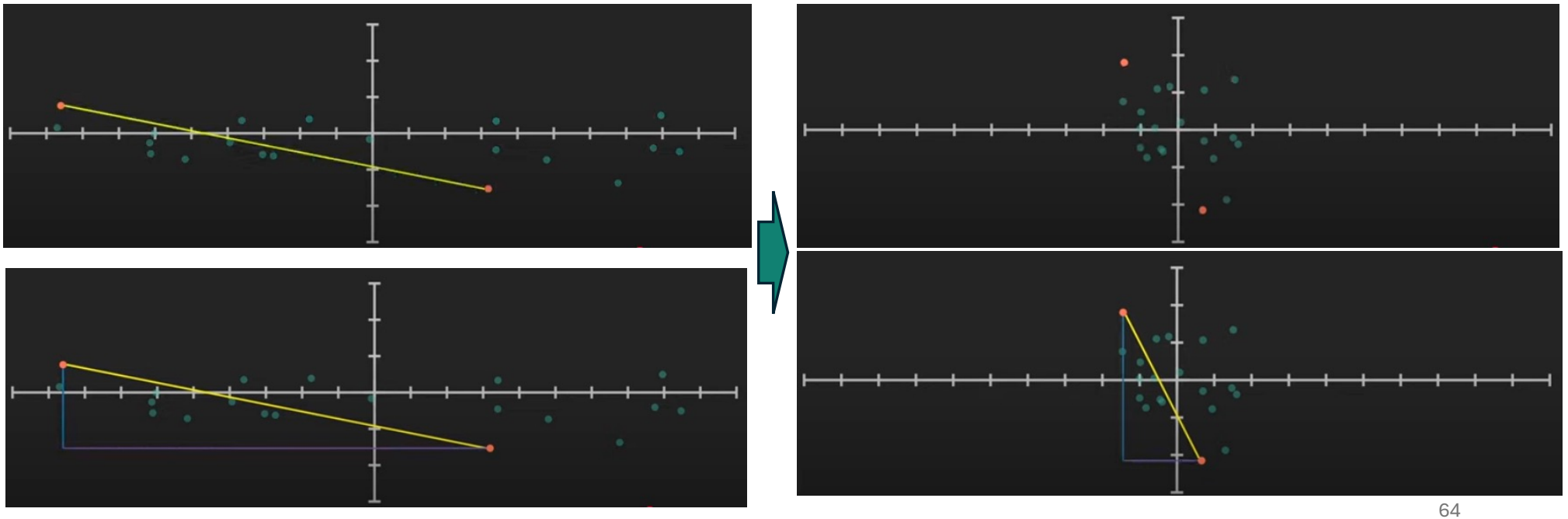
- **Standardization (Z-score Scaling):**
  - Here, each feature is transformed to have a mean of 0 and a standard deviation of 1.
  - This is achieved by subtracting the mean value and dividing by the standard deviation of the feature.
  - x is the original value you want to standardize, μ (mu) is the mean of the data set, σ (sigma) is the standard deviation of the data set

$$Z = (x - \mu) / \sigma$$

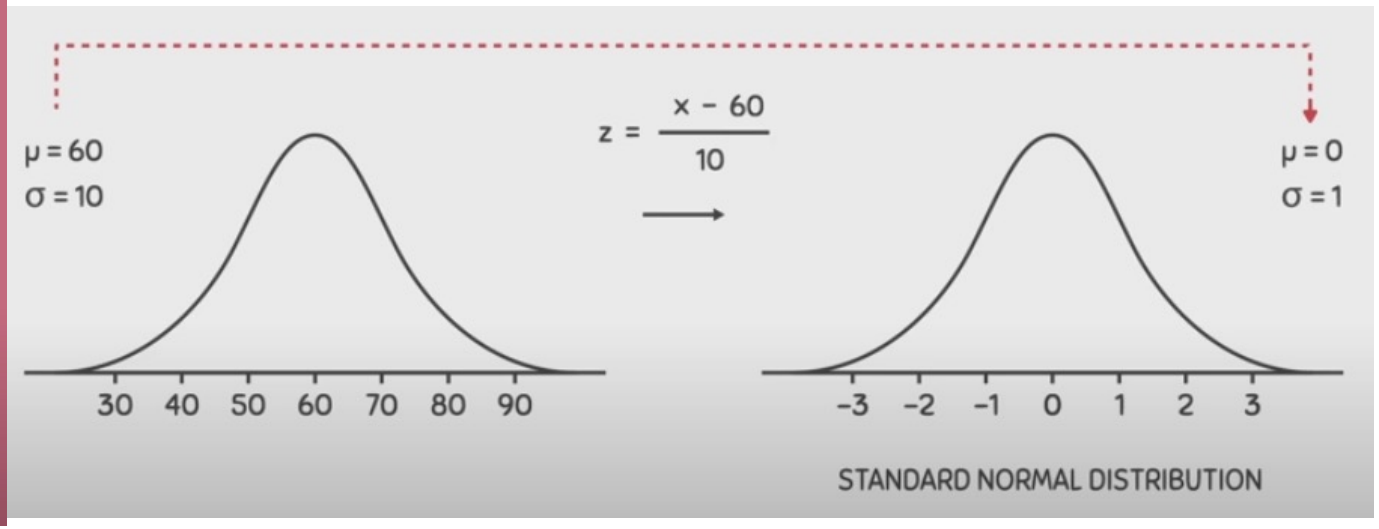
PROGIS
ETUDES, MEDIAS, COMMUNICATION, MARKETING

Feature Scaling

- In the last figure, both feature contribute equally to the distance
- After that, your algorithm won't be affected by the feature with a higher scale.



```
from sklearn.preprocessing import StandardScaler
stdsc=StandardScaler()
X_std=pd.DataFrame(stdsc.fit_transform(X_all), index=X_all.index, columns=X_all.columns)
```



Any normal distribution with any value of mean and standard deviation(sigma) can be transformed into the standard normal distribution where we have a mean of zero and a standars deviation of one.

END