MODULE 03

UN-SUPERVISED LEARNING

WHAT IS UN-SUPERVISED LEARNING ?

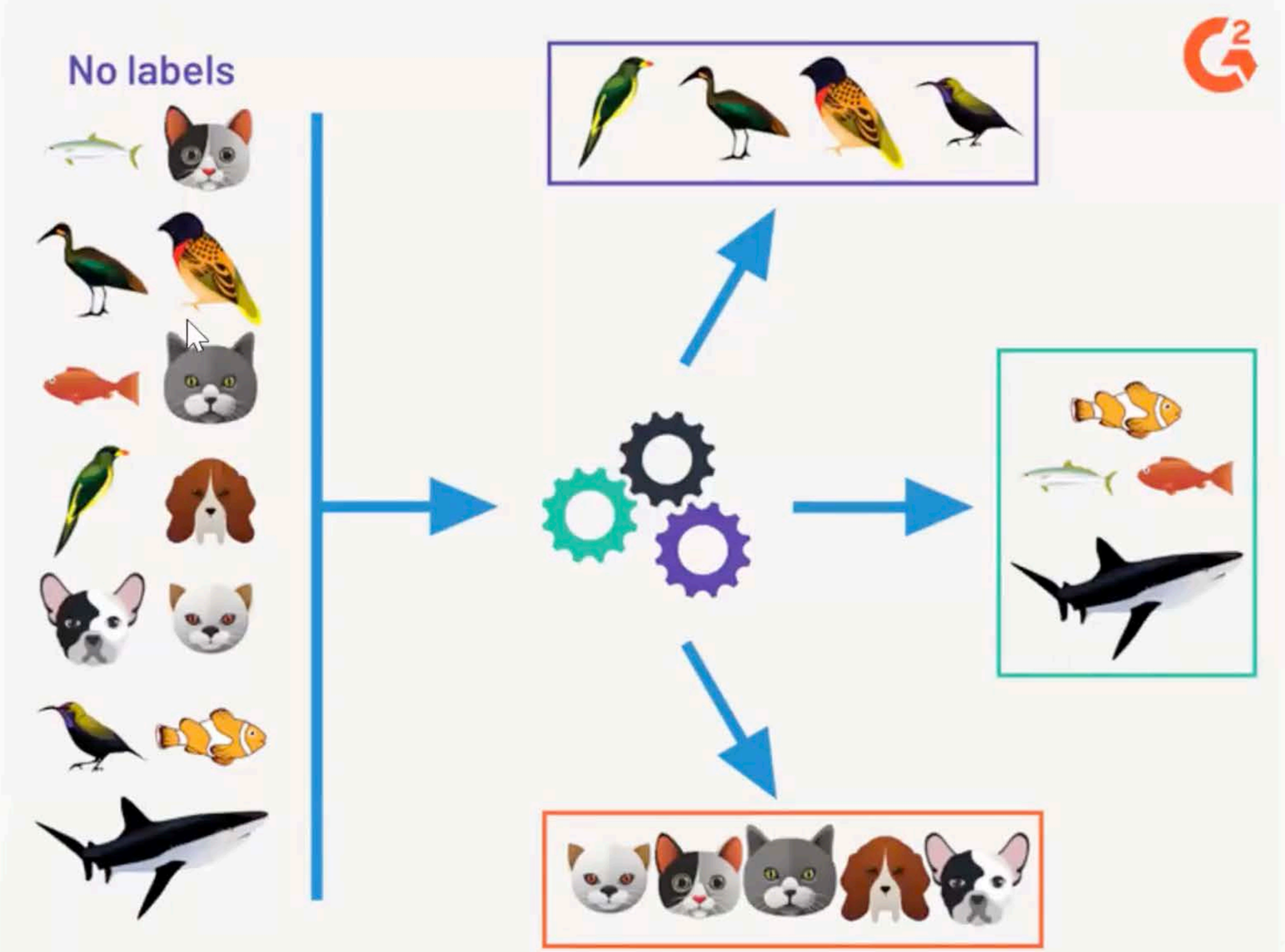
- Unsupervised learning allows us to approach problems with little or no idea what our results should look like that is you only have input data (X) and no corresponding output variables.
- We can derive structure from data where we don't necessarily know the effect of the variables.

GOAL OF USING UN-SUPERVISED LEARNING

- The goal for unsupervised learning is to model the underlying structure or distribution in the data in order to learn more about the data and all this is done without any supervision or direction unlike supervised learning
- The model is given a dataset which is neither labelled nor classified. The model explores the data and draws inferences from datasets to define hidden structures from unlabelled data

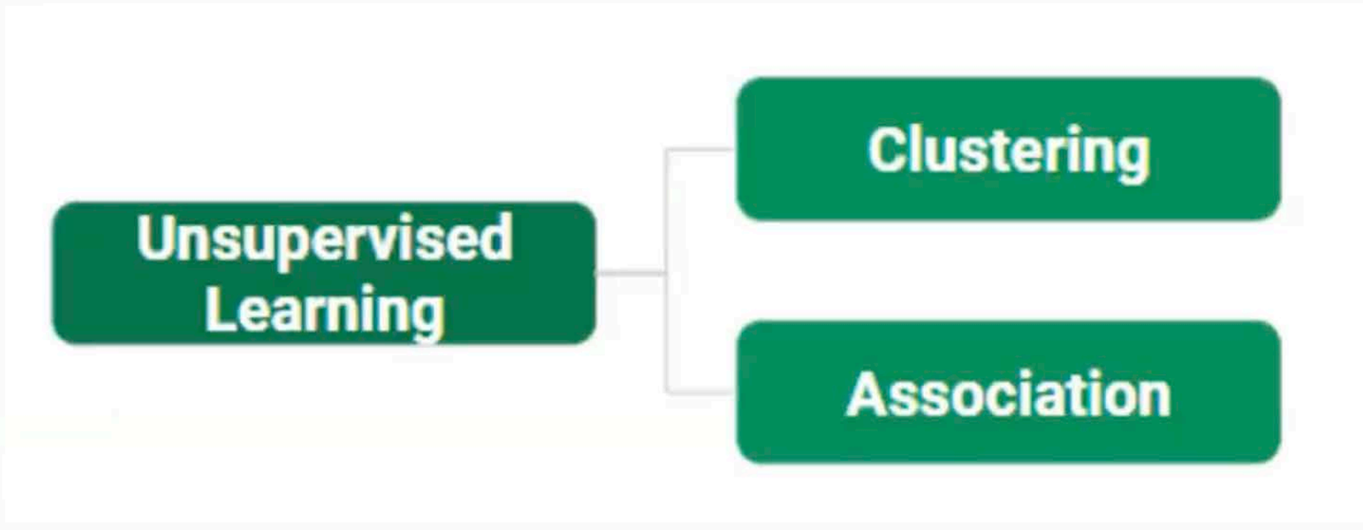
EXAMPLE OF UN-SUPERVISED LEARNING

we can see that even without labels, the algorithm was able to sort the data based on the structures it identified this is achieved by unsupervised learning



TYPES IN UN-SUPERVISED LEARNING

UN-Supervised learning can be split into two subcategories :
Clustering and association

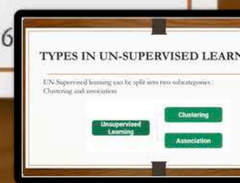


CLUSTERING:

A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behavior.

ASSOCIATION:

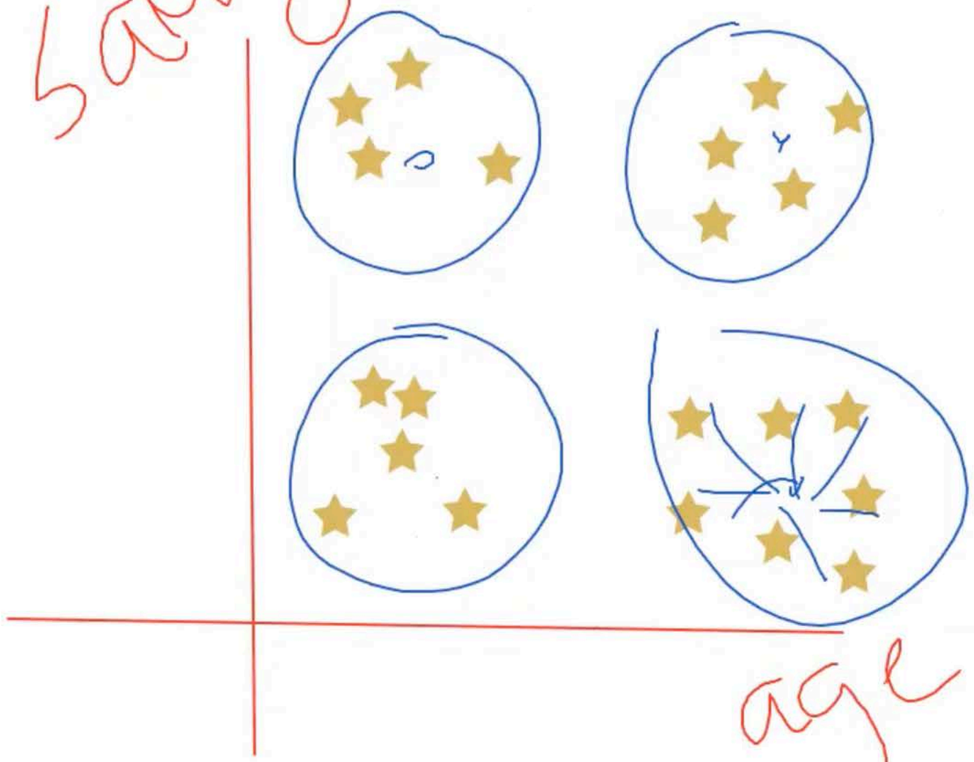
An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.



APPLICATIONS OF UN-SUPERVISED LEARNING

- Customer Segmentation
- Search Engines
- identification of cancer cells
- Handwritten digit recognition
- Clustering used in Search engines

Q → How many clusters?

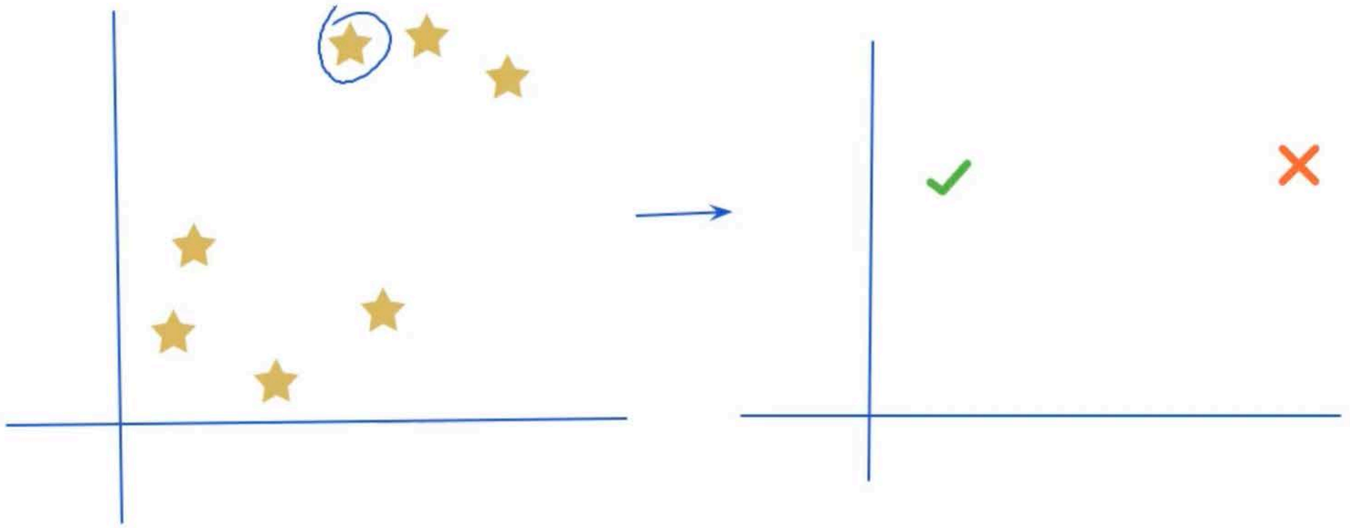


n = 1, 2, 3, 4, ——— 10

Inertia

Elbow Method

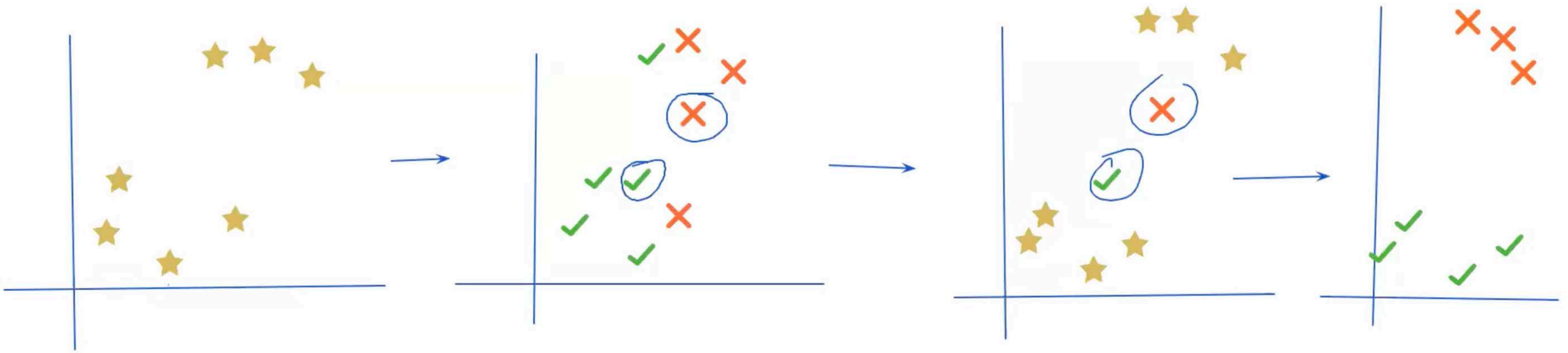
WCSS



Suppose $n(=2)$

2. Take $n(=2)$ pts anywhere

3. For every pt, find which category pt is closer



2. Suppose n(=2)
3. For every pt, find which category pt is closer
4. Take centroid of the new clusters
5. Repeat step 3 & 4