Unsupervised Learning

-Definition

- We don’t teach the model or load it with info. i.e We don’t provide it with labeled dataset.
- The model works on its own to discover info that may not be visible to humans, i.e. we use algorithms that draw conclusions on unlabeled data
    - ⇨ It’s more difficult than supervised:
        Little to know info about the data and the expected output.

- It has fewer tests and models that can be used to ensure that the outcome of the model is accurate => Less controllable environment as the machine is creating outcomes for us.

- This model relies on Clustering: Analysis of patterns and groupings of unlabeled data. We don’t interfere when the algorithm is learning

Example: Watching a match of a sport you don’t know with your friend without them teaching you anything

Clustering has been widely used across industries for years:

- Biology - for genetic and species grouping;
- Medical imaging - for distinguishing between different kinds of tissues;
- Market research - for differentiating groups of customers based on some attributes
- Recommender systems - giving you better Amazon purchase suggestions or Netflix movie matches.

-Association rule (Apriori Algorith):

Supermarket example

Relationships of products(items) in different sets, and the frequency of association. Eliminating the association below a certain threshold. Or to find the one highest (Max.) association between products

Get frequent items or frequent item sets

- Support measure shows how popular the item is by the proportion of transaction in which it appears.
- Confidence measure shows the likeness of Item B being purchased after item A is acquired.
- Lift measure also shows the likeness of Item B being purchased after item A is bought. However, it adds to the equation the demand rate of Item B.