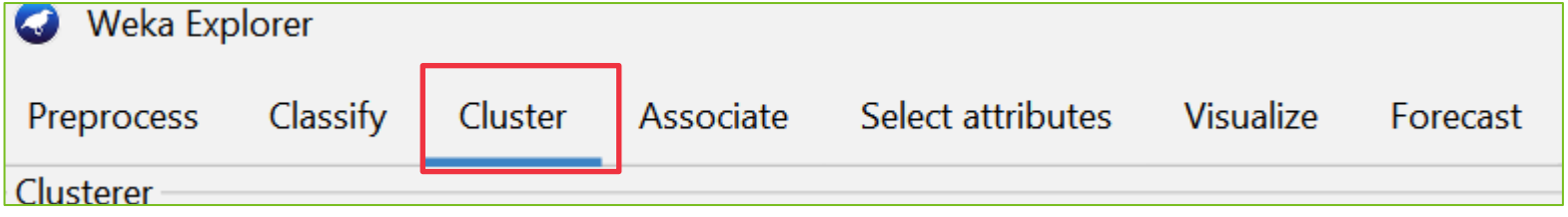
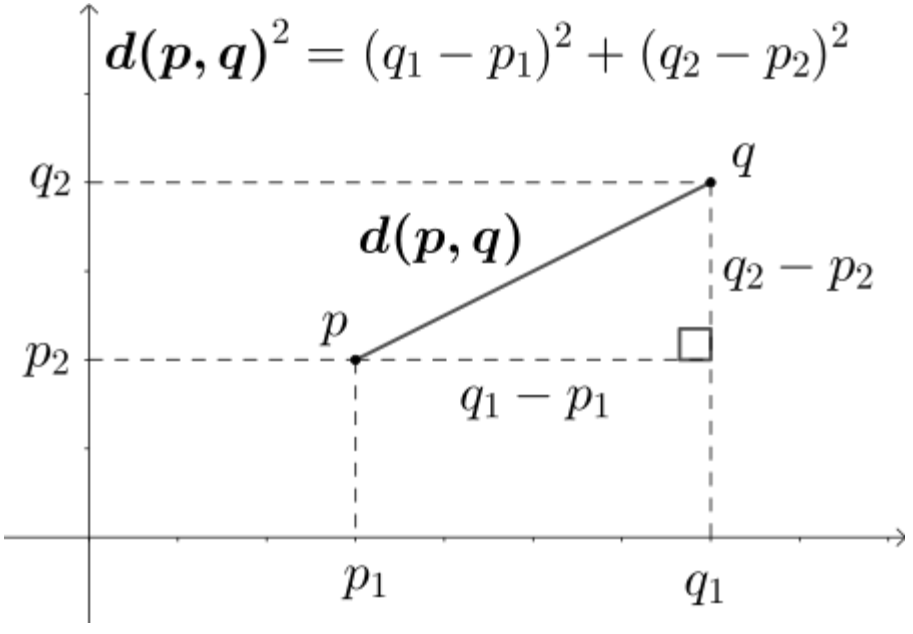
Outline

- **Euclidean Distance**
- **Cluster analysis**
- **Customer segmentation**
- **How to implement cluster analysis in Weka**



Euclidean Distance

ID	p	q	p2
1	3	5	4
2	4	4	5
3	2	5	4



https://www.cuemath.com/euclidean-distance-formula/

Higher dimensions

In three dimensions, the distance is

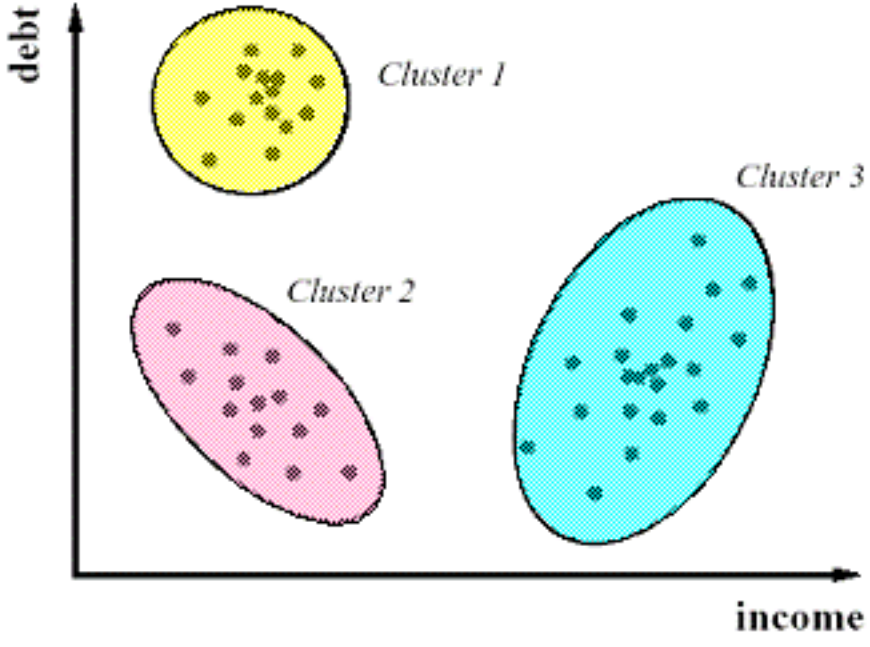
$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2}.$$

In *N* dimensions, the distance is

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_i - q_i)^2 + \cdots + (p_n - q_n)^2}.$$

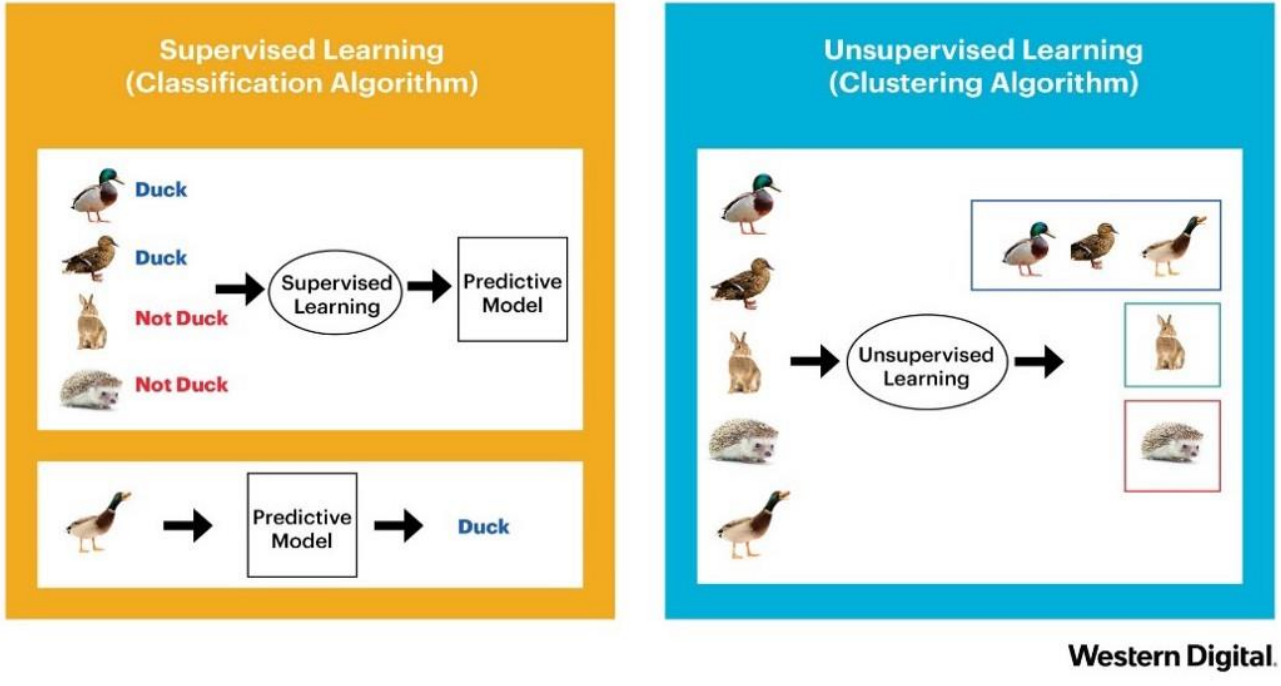
https://en.wikipedia.org/wiki/Euclidean_distance

Cluster



https://www.analyticsvidhya.com/blog/2013/11/getting-clustering-right/

Supervised vs Unsupervised Learning



https://medium.com/hengky-sanjaya-blog/supervised-vs-unsupervised-learning-aae0eb8c4878

Customer segmentation

Customer segmentation is the practice of dividing a company's customers into groups that reflect similarity among customers in each group. The goal of segmenting customers is to decide how to relate to customers in each segment in order to maximize the value of each customer to the business.

https://www.optimove.com/resources/learning-center/customer-segmentation
https://www.qualitygurus.com/types-of-customers-and-customer-segmentation/