Euclidean Distances

Euclidean distance is a measure of the distance b/w two points in multi-dimensional space.

It is calculated as the square root of the sum of the squares of the differences between the coordinates of the two points.

The Importance of Euclidean distance in machine learning is that it is often used as a similarity metric between two data points.

This can be used in clustering algorithms, for example, where data points that are closer together of Euclidean distance are more likely to be grouped togther in the same cluster.

Addionally, it is used in K-nearest neighbors algorithm and linear regression model.

We will calculate the Euclidean distance of one object and display its Shillhouette Score, in order to illustrate the math behind

`euclidean_distances` behind `Sklearn.metrics`

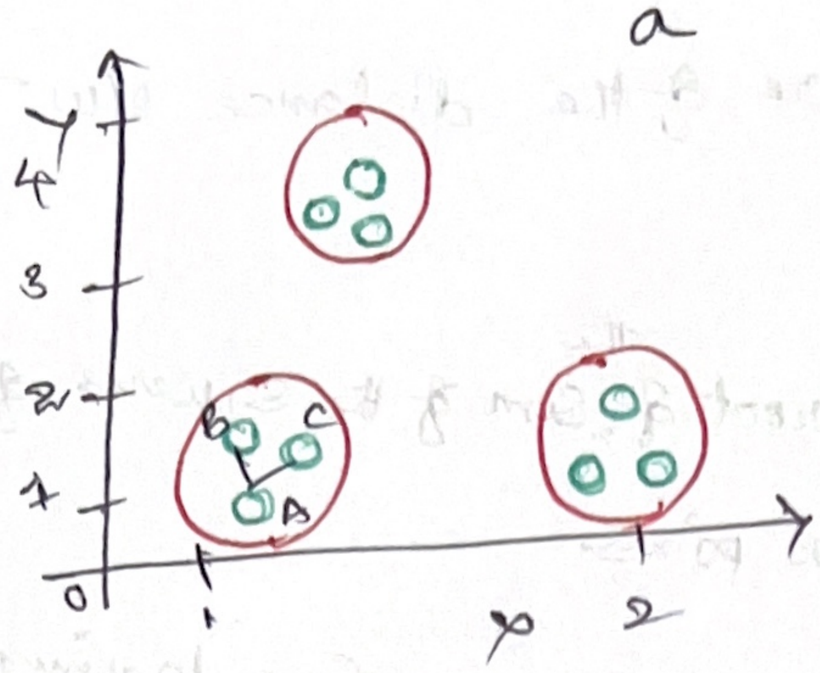
for the purposes of exemplification, we will only calculate the distance between one object within the cluster and the other ones objects that belong cluster and also the distance between the aforementioned point and other elements in the other clusters and the Silhoutte Score of the object.

$$S = \frac{\beta - \alpha}{max(\alpha, \beta)}$$

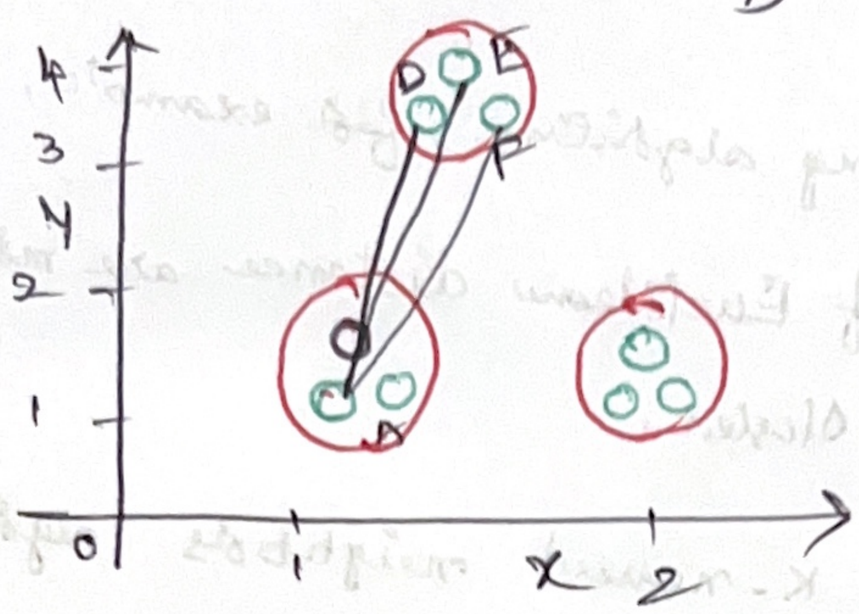
α (alpha) = the average distance between our element and the others within the cluster.

β (beta) = the distance between our element and the closest cluster

a



b



$$d(A,B) = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2}$$

$$d(A,B) = \sqrt{(1-1)^2 + (1,7-0,9)^2}$$

→ We will use object oriented programming. After creating the class Euclid

We will instaliate an object receiving the parameters a, b, c, d, e, f, g

k and i.

→ These parameters will be names randomly generated by names.getfir

and the distances will be arbitary, given the purpose of showing how

to calculate the Silhouette score.

→ The function clusters will return three lists, containing the elements

in each cluster and A.

✓ The function cluster_a will return the Euclian distances between A-B and A-C as well as self.alpha and a DataFrame with the elements of cluster and its respectives distances to x and y

✓ The function cluster_b will return the Euclian distances b/w A-D, A-E & A-F, as well as the avg distance b/w A-cluster 2 and a DataFrame with the elements of cluster and its respective distances to x and y

✓ The function cluster_c will return the Euclidan distance b/w A-G, A-H & A-I, as well as the average distance b/w A-cluster 3 and a DataFrame with the elements of cluster and its respective distances to x and y

✓ The function silhouette will return the values of β (beta), α (alpha) and the Silhoute score of A.

✓ The silhouette of a cluster is the average of the score of all of its elements.