**K-means clustering**

K-means clustering is an unsupervised machine learning algorithm which groups data points within an Eucledian plane (bidimensional or higher) into clusters based on the distances between each point on that plane (and therefore the underlying similarities in the characteristics represented in that plane; Russel and Norvig, 2021).

The algorithm generally follows a series of 3 steps:

1. Initialisation

First, a number of k data points are chosen (by the user, randomly, or using mathematical approaches) as the initial centroids for each cluster. The optimal number of k can be chosen using multiple approaches.

1. Reassign points

In this step, each point is assigned to the cluster for which the distance in the Euclidean plane between its centroid and that data point is shortest.

1. Update centroids

Once all data points have been allocated to a centroid and respective cluster, the location of each centroid is updated so it is now located at central location of the cluster (i.e. the position which minimises the distances between the centroid and all data points assigned to that cluster).

Steps 2 and 3 are then repeated sequentially until the algorithm converges – i.e. no further changes are made to cluster allocations.

K-means are a commonly used algorithm for clustering purposes, as they are computationally fast, simple to implement, and easily comprehensible and interpretable (as well illustrated in the two suggested step-by-step visualisations; Harris, 2014; Shabalin, ND). However, they also highlight some of their limitations, namely the impact of initialisation process and centroids, as well as the choice of k, in the resulting clusters, together with the possible variability in the resulting clusters even when using the same random initialisation strategy, and their challenges in handling data spaces which are not regularly shaped or equally sized.

**References:**

Harris, N. (2014) *Visualizing K-Means Clustering*. Available from: https://www.naftaliharris.com/blog/visualizing-k-means-clustering/ (Accessed: 10 December 2024).

Russel, S. & Norvig, P. (2021) *Artificial intelligence: A modern approach*. Global Edition. Pearson Higher Education.

Shabalin, A. (ND) *K-means clustering*. Available from: https://shabal.in/visuals/kmeans/1.html (Accessed: 10 December 2024).