

k centroids (quasi-data points representing the centres of the clusters) are distributed at random among the data points

The distance (usually Euclidean distance, but other measures can be used) between every data point and each centroid is calculated

Each data point is associated with the nearest centroid

The centroids move to minimise the distance between them and their associated points, so moving to the centre of their points

Have the centroids moved a sufficiently small amount – i.e. has convergence been reached?

No

Yes

The configuration of clusters is remembered

Has the limit on the number of iterations been reached?

No

Yes

The cluster configuration with the smallest distance between points and their associated centroids is output as the clustering solution

Finish