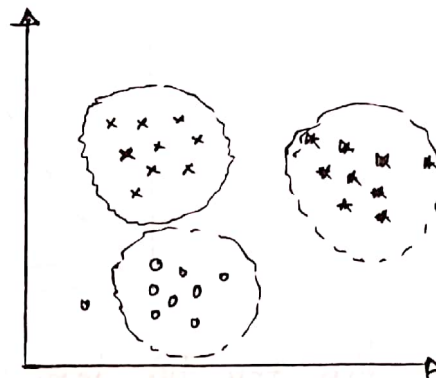


K-Means Clustering



Goal! Cluster a dataset into k different clusters. The dataset is unlabeled (unsupervised learning). Each sample is assigned to the cluster with the nearest mean.

Iterative Optimization!

1. Initialize cluster centers (e.g. randomly)

2. Repeat until converged:

→ Update cluster labels: Assign points to the nearest cluster center (centroid)

→ Update cluster centers (centroids): set center to the mean of each cluster.

Euclidean Distance!

Get the distance between two feature vectors

$$d(p, q) = \sqrt{\sum (p_i - q_i)^2}$$