

K-Means Clustering Algorithm

Problem Statement: given a two-dimensional dataset X , create two clusters.

Initialize 2 centroids, e.g., randomly take 2 instances from X

$$\text{centroid 0} = x^{(i)}, \text{centroid 1} = x^{(j)}$$

Repeat until convergence (e.g., no change in cluster assignments or centroids) {

- Compute the Euclidean distance between every instance $x^{(k)}$ and centroids 0 & 1

Assign instance $x^{(k)}$ to cluster 0 if it is closer to centroid 0

Assign instance $x^{(k)}$ to cluster 1 if it is closer to centroid 1

- Recompute centroids 0 and 1

centroid 0 = the average of all instances assigned to cluster 0

centroid 1 = the average of all instances assigned to cluster 1

}

Recall: Euclidean distance between $a^T = [a_1, a_2]$ and $b^T = [b_1, b_2]$ is $\sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}$