K-means Clustering

- allows us to cluster unlabeled data in an unsupervised learning algorithm that will attempt to group similar clusters together in your data.

- Typical clustering problem look like?
  - ↳ Cluster similar documents
  - ↳ Cluster customers based on features
  - ↳ Market Segmentation
  - ↳ Identify similar physical groups

- The overall goal is to divide data into distinct groups. Such that observations within each group are similar.

Algorithm

Example:

Height	weight
185	72
170	56
168	60
179	68
182	72
188	77
180	71
183	84
180	67

1. Assume first two rows to form a cluster K1 and K2.

2. Each cluster has its central value i.e centroid

3. on the basis of central value, you can group up remaining atoms.
   - First row as a centroid of K1 → (185,72)
   - Second row as a centroid of K2 → (170,56)

3. Now, we see in which remaining data belongs to. by using Euclidean Distance

$$K = \sqrt{(x_o - x_c)^2 + (y_o - y_c)^2}$$ observed - centroid

K1 (185,72)   K2 (170,56)
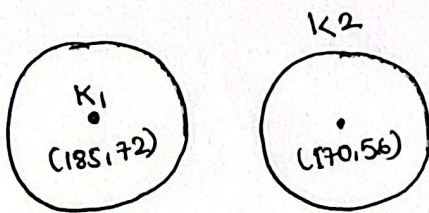
$$K_1 = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} = \sqrt{(168 - 185)^2 + (60 - 72)^2} = 20.80$$

$$K_2 = \sqrt{(168 - 170)^2 + (60 - 56)^2} = 4.48$$

The distance between observed value and centroid of K1, K2 is 4·48 & 20·80.

Select one with less distance i.e 4·48 which is K_2 cluster.

4. If one row falls under a cluster, you need to create a new centroid. Here are need to compute new centroid. So,

$$K_2 = \left(\frac{x_2 + x_1}{2}, \frac{y_2 + y_1}{2}\right) = \left(\frac{170 + 168}{2}, \frac{170 + 56 + 60}{2}\right) = (169, 58)$$

5. Now, centroid of K_2 becomes $(169, 58)$ & $K_1 (185, 72)$.

6. Continue computing with following data.

(II). For $(179, 68)$: Observed data.

$$K_1 = \sqrt{(179 - 185)^2 + (68 - 72)^2} = \sqrt{36 + 16} = \sqrt{52} = 7·2$$

$$K_2 = \sqrt{(179 - 169)^2 + (68 - 58)^2} = \sqrt{100 + 100} = \sqrt{200} = 14·14$$

Here (179, 68) falls under K_1 cluster so, the K_1 cluster is to be recomputed again.

$$K_1 = \left(\frac{185 + 179}{2}, \frac{68 + 72}{2}\right) = (182, 70)$$

updated: $K_1 = (182, 70)$ $K_2 = (169, 58)$ ✓