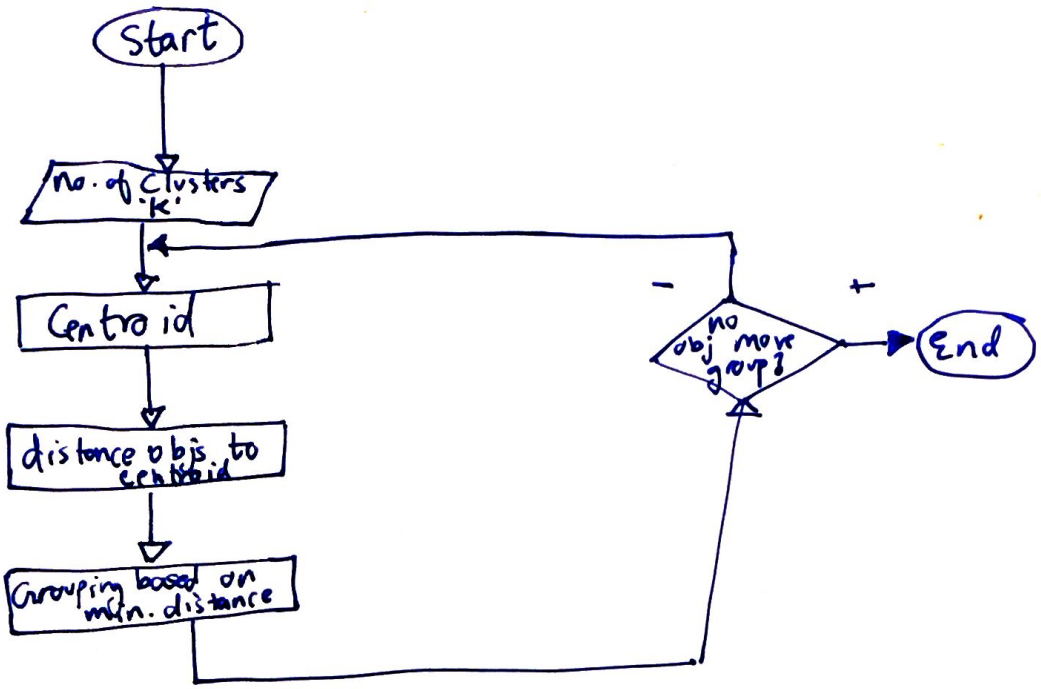
K - means Clustering:



→ define no. of clusters k,

→ Centroid = mean value, 2 → 2 centroids
3 → 3 centroids

→ find dist. of obj from centroids.

→ Based on $\overset{\text{min}}{\wedge}$ distance, we classify to a specific cluster

→ repeat process: if group changes, we move to cluster diff

Example
for K=3

Data: 2 5 7 8 13 16 18 28 30

C_1 assume (2) C_2 assume (13) C_3 assume (30)

1) C_1 [2 5 7] M = 4.6 C_2 [8 13 16 18] M = 13.5 [28 30] M = 29

2) [2 5 7 8] M = 5.5 [13 16 18] M = 15.6 [28 30] M = 29 — run loop again — find centroids

loop ends

K-means in 2 features:

using K=2.

Initiation - Randomly we chose following (two Centroids for 2 clusters)

I) In this case, the two centroids are:

$$m_1 = (1.0, 1.0) \text{ \& } m_2 = (5.0, 7.0)$$

Data

Individual	Var$_1$	Var$_2$
1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5

	Indiv.	μ Vector
Group 1	1	(1.0, 1.0)
Group 2	4	(5.0, 7.0)

$$\sqrt{(m-x)^2 + (m-y)^2}$$

x: C_1, y: C_2

m value which we use to calc. value wrt

eg

$$d(m_1, 2) = \sqrt{|1.0-1.5|^2 + |1.0-2.0|^2} = 1.12$$

$$d(m_2, 2) = \sqrt{|5.0-1.5|^2 + |7.0-2.0|^2} = 6.10$$

II) → we do this for all observations. (& create an internal table of sorts)

→ we thus get 2 clusters containing {1, 2, 3} & {4, 5, 6, 7}

their new centroids are

	Cent 1	Cent 2
1	0	7.21
2	1.12	6.10
3	3.61	3.61
4	7.21	0
5	4.72	2.5
6	5.30	2.06
7	4.30	2.92

$$m_1 = \left(\frac{1}{3}(1.0+1.5+3.0), \frac{1}{3}(1.0+2.0+4.0)\right) = (1.83, 2.33)$$

$$m_2 = \left(\frac{1}{4}(5.0+3.5+4.5+3.5), \frac{1}{4}(7.0+5.0+5.0+4.5)\right) = (4.12, 5.38)$$

III) - Now using these centroids, we compute the Euclidean distance of each object as shown

- New clusters are {1, 2} & {3, 4, 5, 6, 7}
- Next centroids are $m_1 = (1.25, 1.5)$ & $m_2 = (3.9, 5.1)$

	Cent 1	Cent 2	
1	1.57	5.38	Result
2	0.47	4.28	
3	2.04	1.78	
4	5.64	1.84	
5	3.15	0.73	
6	3.78	0.54	
7	2.74	1.08	

IV) we run again until observations stop changing groups. (we're calculating clusters again after each step)