

$\{1, 2, 5, 6, 9, 10\}$

C

Initial Seeds: $S'_1 = 1$ $S'_2 = 2$ $S'_3 = 5$
 $\{1\}$ $\{2\}$ $\{5, 6, 9, 10\}$

New Seeds: $S'_1 = 1$ $S'_2 = 2$ $S'_3 = 7.5$
 $\{1\}$ $\{2\}$ $\{5, 6, 9, 10\}$

New Seeds: $S'_1 = 1$ $S'_2 = 2$ $S'_3 = 7.5$

$$\begin{aligned} ESS &= (1-1)^2 + (2-2)^2 + (5-7.5)^2 + (6-7.5)^2 + (9-7.5)^2 \\ &\quad + (10-7.5)^2 \\ &= 17 \end{aligned}$$

D

Initial Seeds: $S'_1 = 6$ $S'_2 = 9$ $S'_3 = 10$
 $\{1, 2, 5, 6\}$ $\{9\}$ $\{10\}$

New Seeds: $S'_1 = 3.5$ $S'_2 = 9$ $S'_3 = 10$
 $\{1, 2, 5, 6\}$ $\{9\}$ $\{10\}$

New Seeds: $S'_1 = 3.5$ $S'_2 = 9$ $S'_3 = 10$

$$\begin{aligned} ESS &= (1-3.5)^2 + (2-3.5)^2 + (5-3.5)^2 + (6-3.5)^2 + (9-9)^2 + (10-10)^2 \\ &= 17 \end{aligned}$$

Intuitively

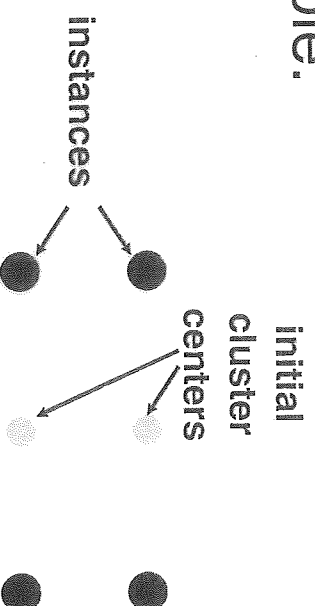
Works with numeric data only

- 2) Pick a number (k) of cluster centers (at random)
- 3) Assign every item to its nearest cluster center (e.g., using Euclidean distance)
- 4) Move each cluster center to the mean of its assigned items
- 5) Repeat steps 2 and 3 until convergence (i.e., change in cluster assignments less than a threshold)

Discussion

- Result can vary significantly depending on initial choice of seeds
- Can get trapped in local minimum

– Example:



- To increase chances of finding global optimum: restart with different random seeds