Lab -09

Build K-means algorithm to cluster a set of data stored in a .csv file.

Compute two clusters using k-means for clustering when initial cluster centers are (1,1) & (5,7). Execute for two iterations.

→ Iteration 1:
  Calculate euclidian distance b/w
         (1,1)   (5,7)
  Reroanch   close to C1   Close to C2   Assignment

Point		close to C1	Close to C2	Assignment
(1,1)		0	7.1	C1
(1.5,2)		1.12	6.12	C1
(3,4)		3.61	3.61	C1
(5,7)		7.21	0.0	C2
(3.5,5)		4.12	2.5	C2
(4.5,5)		5.31	2.06	C2
(3.5,4.5)		4.30	2.92	C2

New centroids:

$$C1 = \frac{1+1.5+3}{3}, \frac{1+2+1}{3} = 1.83, 2.33$$

$$C2 = \frac{5+3.5+4.5+3.5}{4}, \frac{7+5+5+4.5}{4} = 4.12, 5.32$$

Iteration · 2

	(1·83, 2·33)	(4·17, 5·37)	
Record	close to C_1	close to C_2	Assignment
(1,1)	1·67	5·37	C_1
(1·5, 2)	0·47	4·27	C_1
(3, 4)	2·04	1·77	C_2
(5, 7)	5·64	1·85	C_2
(3·5, 5)	3·15	0·72	C_2
(4·5, 5)	3·78	0·53	C_2
(3·5, 4·5)	2·74	1·07	C_2

New cluster one

$C_1 = R_1, R_7$
$C_2 = R_3, R_4, R_5, R_6$

New centroids

$C_1 = \dfrac{25}{2}, \dfrac{3}{2}$

$C_2 = \dfrac{19.5}{5}, \dfrac{25.5}{5}$

For iris.csv Dataset

The elbow plot criteria vector shows a sharp elbow at $k=3$, indicating that the cluster is the optimal choice for the sepal length/width of the iris.