Question 1: Solving k-Means by hand

$P_1(10,20)$, $P_2(50,10)$, $P_3(55,15)$, $P_4(15,15)$, $P_5(20,10)$

Assume $k=2$ and the initial centroids are chosen as
$C_1 = (15,15)$, $C_2 = (50,10)$

(1) Compute the Euclidean distance of each point from both centroids

(*) $P_1(10,20)$

Distance to $C_1 = \sqrt{(10-15)^2 + (20-15)^2} = 7.07$

Distance to $C_2 = \sqrt{(10-50)^2 + (20-10)^2} = 41.23$

(*) $P_2(50,10)$

Distance to $C_1 = \sqrt{(50-15)^2 + (10-15)^2} = 35.35$

to $C_2 = \sqrt{(50-50)^2 + (10-10)^2} = 0$

(*) $P_3(55,15)$

Distance to $C_1 = \sqrt{(55-15)^2 + (15-15)^2} = 40$

to $C_2 = \sqrt{(55-50)^2 + (15-10)^2} = 7.07$

(*) $P_4(15,15)$

Distance to $C_1 = \sqrt{(15-15)^2 + (15-15)^2} = 0$

to $C_2 = \sqrt{(15-50)^2 + (15-10)^2} = 35.35$

(*) $P_5(20,10)$

Distance to $C_1 = \sqrt{(20-15)^2 + (10-15)^2} = 7.07$

to $C_2 = \sqrt{(20-50)^2 + (10-10)^2} = 30$

(2) Assign each point to the closest cluster

$P_1 (10,20) \Rightarrow$ assign : C_1

$P_2 (50,10) \Rightarrow$ assign : C_2

$P_3 (55,15) \Rightarrow$ assign : C_2

$P_4 (15,15) \Rightarrow$ assign : C_1

$P_5 (20,10) \Rightarrow$ assign : C_1

(3) Compute the new centroids by averaging the points in each cluster

First iteration, update step

The mean (average) of data assigned to centroid C_1:

$$C_1.x = \frac{10+15+20}{3} = 15$$

$$C_1.y = \frac{20+15+10}{3} = 15$$

The mean of data assigned to centroid C_2:

$$C_2.x = \frac{50+55}{2} = 52.5$$

$$C_2.y = \frac{10+15}{2} = 12.5$$

Second iteration, assignment step:

	Distance to C_1 (15, 15)	Distance to C_2 (52.5, 12.5)	Assign
P_1 (10,20)	7.07	43.16	C_1
P_2 (50,10)	35.35	3.54	C_2
P_3 (55,15)	40	3.54	C_2
P_4 (15,15)	0	37.58	C_1
P_5 (20,10)	7.07	32.6	C_1

Second iteration, update step:

The assignments did not change and therefore the means are the same. The algorithm has converged.

Two centroids are: $C_1(1.5, 1.5)$ and $C_2(52.5, 12.5)$