K- MEDOID

cluster the given data using K- Medoid technique.
Use K = 2

{(2,6), (3,4), (3,8), (4,7), (6,2), (6,4), (7,3), (7,4), (8,5) (7,6)}

solution:

Using Manhatten / City Block distance measure to calculate the dissimilarity of the data objects.

$$d = |x_1 - x_2| + |y_1 - y_2|$$

X	Y	$c_1 = (3,4)$ $d = \|x-3\| + \|y-4\|$	$c_2 = (7,4)$ $d = \|x-7\| + \|y-4\|$	C
2	6	$\|2-3\| + \|6-4\| = 3$	$\|2-7\| + \|6-4\| = 7$	c_1
3	4	$0+0 = 0$	$4+0 = 4$	c_1
3	8	$0+4 = 4$	$4+4 = 8$	c_1
4	7	$1+3 = 4$	$3+3 = 6$	c_1
6	2	$3+2 = 5$	$1+2 = 3$	c_2
6	4	$3+0 = 3$	$1+0 = 1$	c_2
7	3	$4+1 = 5$	$0+1 = 1$	c_2
7	4	$4+0 = 4$	$0+0 = 0$	c_2
8	5	$5+1 = 6$	$1+1 = 2$	c_2
7	6	$4+2 = 6$	$0+2 = 2$	c_2

Initial Medoids = {(3,4), (7,4)}

Generated clusters based on distance values from each medoid

$c_1 = \{(2,6), (3,4), (3,8), (4,7)\}$

$c_2 = \{(6,2), (6,4), (7,3), (7,4), (8,5), (7,6)\}$

Let's compute the cost for this iteration by adding the distance values.

Cost - Iteration 1 =
$(3+0+4+4) + (3+1+1+0+2+2)$
$= 20$

Iteration 2: select one of the non-medoid data object

Let new medoid be (7,3)

X	Y	$C_1=(3,4)$ $d=\|x-3\|+\|y-4\|$	$C_2=(7,3)$ $d=\|x-7\|+\|y-3\|$	C
2	6	$\|2-3\|+\|6-4\|=3$	$\|2-7\|+\|6-3\|=8$	C_1
3	4	$0+0=0$	$4+1=5$	C_1
3	8	$0+4=4$	$4+5=9$	C_1
4	7	$1+3=4$	$3+4=7$	C_1
6	2	$3+2=5$	$1+1=2$	C_2
6	4	$3+0=3$	$1+1=2$	C_2
7	3	$4+1=5$	$0+0=0$	C_2
7	4	$4+0=4$	$0+1=1$	C_2
8	5	$5+1=6$	$1+2=3$	C_2
7	6	$4+2=6$	$0+3=3$	C_2

Generating clusters based on distance from newly chosen Medoid

$C_1 = \{(2,6), (3,4), (3,8), (4,7)\}$

$C_2 = \{(6,2), (6,4), (7,3), (7,4), (8,5), (7,6)\}$

Let's compute the cost for this iteration by adding their corresponding distance values.

$$\text{cost\_Iteration2} = (3+0+4+4)+(2+2+0+1+3+3)$$
$$= 11+11$$
$$= 22$$

Step 3: So cost of swapping medoid from I_1 & I_2

$$= \text{cost\_Iteration2} - \text{cost\_iteration1}$$
$$= 22 - 20$$
$$= 2$$

Swap cost is 2 which is >0, so the medoid chosen in iteration 2 is not a good idea.

So we will go with the chosen medoids in Iteration 1.