

Homework 5: Clustering

กำหนดให้ training dataset $X = \{ [2, 8], [1, 6], [1, 8], [2, 7], [8, 2], [6, 1], [8, 1], [7, 2] \}$ ให้ใช้ k-means algorithm ในการคำนวณ cluster centroids โดยกำหนดค่า $K = 2$ และ cluster centroid เริ่มต้น $\mu_1 = [3, 6]$ และ $\mu_2 = [6, 3]$

```
Repeat{
  #Assign Points to Cluster Centroids
  for i = 1 to m:
     $c^{(i)}$  = index(from 1 to K) of cluster centroids closet to  $x^{(i)}$ 

  #Move Cluster Centroid
  for k = 1 to K:
     $\mu_k$  = average (mean) of points assign to cluster k
}
```

© Ter

Step 1 #Assign Points to Cluster Centroids

1.1: คำนวณหา distance ระหว่าง training dataset X และ μ_1, μ_2 ด้วยสูตร $\|x^{(i)} - \mu_k\|$

$$\|x^{(1)} - \mu_1\| = \sqrt{(2-3)^2 + (8-6)^2} = \sqrt{5} \quad \|x^{(5)} - \mu_1\| = \sqrt{(8-3)^2 + (2-6)^2} = \sqrt{41}$$

$$\|x^{(1)} - \mu_2\| = \sqrt{(2-6)^2 + (8-3)^2} = \sqrt{41} \quad \|x^{(5)} - \mu_2\| = \sqrt{(8-6)^2 + (2-3)^2} = \sqrt{5}$$

$$\|x^{(2)} - \mu_1\| = \sqrt{(1-3)^2 + (6-6)^2} = \sqrt{4} \quad \|x^{(6)} - \mu_1\| = \sqrt{(6-3)^2 + (1-6)^2} = \sqrt{34}$$

$$\|x^{(2)} - \mu_2\| = \sqrt{(1-6)^2 + (6-3)^2} = \sqrt{34} \quad \|x^{(6)} - \mu_2\| = \sqrt{(6-6)^2 + (1-3)^2} = \sqrt{4}$$

$$\|x^{(3)} - \mu_1\| = \sqrt{(1-3)^2 + (8-6)^2} = \sqrt{8} \quad \|x^{(7)} - \mu_1\| = \sqrt{(8-3)^2 + (1-6)^2} = \sqrt{50}$$

$$\|x^{(3)} - \mu_2\| = \sqrt{(1-6)^2 + (8-3)^2} = \sqrt{50} \quad \|x^{(7)} - \mu_2\| = \sqrt{(8-6)^2 + (1-3)^2} = \sqrt{8}$$

$$\|x^{(4)} - \mu_1\| = \sqrt{(2-3)^2 + (7-6)^2} = \sqrt{2} \quad \|x^{(8)} - \mu_1\| = \sqrt{(7-3)^2 + (2-6)^2} = \sqrt{32}$$

$$\|x^{(4)} - \mu_2\| = \sqrt{(2-6)^2 + (7-3)^2} = \sqrt{32} \quad \|x^{(8)} - \mu_2\| = \sqrt{(7-6)^2 + (2-3)^2} = \sqrt{2}$$

#Assign Points to Cluster Centroids

1.2: หา index หรือ $c^{(i)}$ ของ training dataset แต่ละตัว

i	$x^{(i)}$	k	$\ x^{(i)} - \mu_k\ $	index ($c^{(i)}$) ที่เลือก
1	[2, 8]	1	$\sqrt{5}$	1 หรือ μ_1
	[2, 8]	2	$\sqrt{41}$	
2	[1, 6]	1	$\sqrt{4}$	1 หรือ μ_1
	[1, 6]	2	$\sqrt{34}$	
3	[1, 8]	1	$\sqrt{8}$	1 หรือ μ_1
	[1, 8]	2	$\sqrt{50}$	
4	[2, 7]	1	$\sqrt{2}$	1 หรือ μ_1
	[2, 7]	2	$\sqrt{32}$	
5	[8, 2]	1	$\sqrt{41}$	2 หรือ μ_2
	[8, 2]	2	$\sqrt{5}$	
6	[6, 1]	1	$\sqrt{34}$	2 หรือ μ_2
	[6, 1]	2	$\sqrt{4}$	
7	[8, 1]	1	$\sqrt{50}$	2 หรือ μ_2
	[8, 1]	2	$\sqrt{8}$	
8	[7, 2]	1	$\sqrt{32}$	2 หรือ μ_2
	[7, 2]	2	$\sqrt{2}$	

Step 2 #Move Cluster Centroid

คำนวณหา cluster centroid ใหม่ของ μ_1, μ_2 โดยคำนวณจากค่าเฉลี่ยของ training data ที่อยู่ภายใต้แต่ละ cluster centroid

$$\begin{aligned}\mu_1 &= \text{average} ([2,8], [1, 6], [1, 8], [2, 7]) \\ &= \frac{1}{4} [(2+1+1+2), (8+6+8+7)] \\ \mu_1 &= [1.5, 7.25]\end{aligned}$$

$$\begin{aligned}\mu_2 &= \text{average} ([8,2], [6, 1], [8, 1], [7, 2]) \\ &= \frac{1}{4} [(8+6+8+7), (2+1+1+2)] \\ \mu_2 &= [7.25, 1.5]\end{aligned}$$

ดังนั้น cluster centroid ใหม่ ได้แก่

$$\mu_1 = [1.5, 7.25]$$

และ

$$\mu_2 = [7.25, 1.5]$$