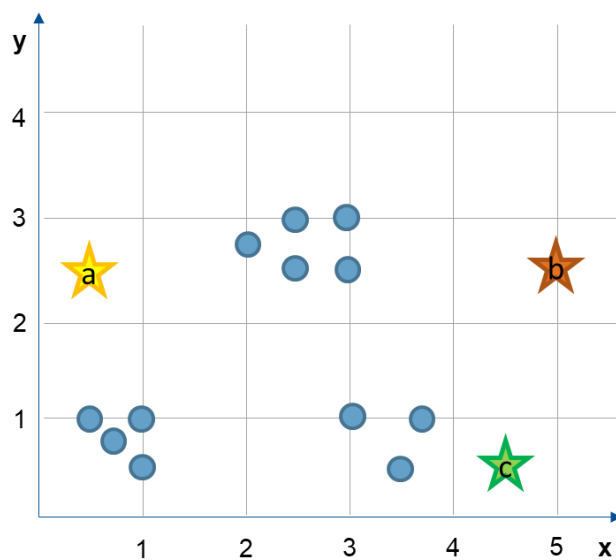# Tutorial Business Analytics

Tutorial 9

Exercise 9.1

Group the data into three clusters applying the k-Means algorithm and the Euclidean distance function.

Dataset				Centroids	
p_i	x	y	c_i	x	y
1	2.5	3		0.5	2.5
2	3	3		5.0	2.5
3	2	2.75		4.5	0.5
4	2.5	2.5			
5	3	2.5			
6	0.5	1			
7	1	1			
8	3	1			
9	3.75	1			
10	0.75	0.75			
11	1	0.5			
12	3.5	0.5			

Solution 9.1

1. Assign instances to nearest cluster centre

p_i	x	y		c_i	x	y
1	2.5	3			0.5	2.5
2	3	3			5.0	2.5
3	2	2.75			4.5	0.5
4	2.5	2.5				
5	3	2.5				
6	0.5	1				
7	1	1				
8	3	1				
9	3.75	1				
10	0.75	0.75				
11	1	0.5				
12	3.5	0.5				

2. Update cluster centre

p_i	x	y		c_i	x	y
1	2.5	3			1.46	1.64
2	3	3			3.00	2.75
3	2	2.75			3.42	0.83
4	2.5	2.5				
5	3	2.5				
6	0.5	1				
7	1	1				
8	3	1				
9	3.75	1				
10	0.75	0.75				
11	1	0.5				
12	3.5	0.5				

3. Assign instances to nearest cluster

p_i	x	y		c_i	x	y
1	2.5	3			1.46	1.64
2	3	3			3.00	2.75
3	2	2.75			3.42	0.83
4	2.5	2.5				
5	3	2.5				
6	0.5	1				
7	1	1				
8	3	1				
9	3.75	1				
10	0.75	0.75				
11	1	0.5				
12	3.5	0.5				

4. Update cluster centre

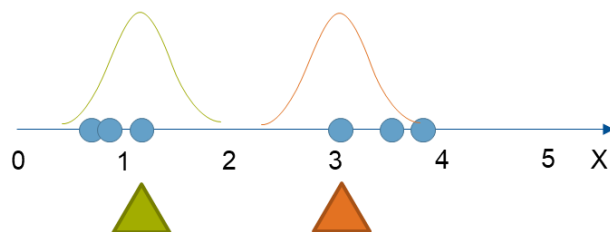| p_i	x	y		c_i	x	y
1	2.5	3			0.81	0.81
2	3	3			2.60	2.75
3	2	2.75			3.42	0.83
4	2.5	2.5				
5	3	2.5				
6	0.5	1				
7	1	1				
8	3	1				
9	3.75	1				
10	0.75	0.75				
11	1	0.5				
12	3.5	0.5				

5. Assign instances to nearest cluster centre

No reassignment → termination

Exercise 9.2

Given k=2, perform EM algorithm with the following instances.

Instance	1	2	3	4	5	6
Value	0.76	0.86	1.12	3.05	3.51	3.75

μ_A	1.12
σ_A	1.00
p_A	50%
μ_B	3.05
σ_B	1.00
p_B	50%

Solution 9.2

Step 1.1: Calculate cluster probabilities

	1	2	3	4	5	6
Pr[A|x]	92.81%	91.41%	86.56%	13.44%	6.01%	3.87%
Pr[B|x]	7.19%	8.59%	13.44%	86.56%	93.99%	96.13%

Step 1.2: Update distribution parameters

μ_A	1.10
σ_A	0.66
p_A	49%
μ_B	3.21
σ_B	0.78
p_B	51%

Step 2.1: Calculate cluster probabilities

	1	2	3	4	5	6
Pr[A|x]	99.25%	98.97%	97.55%	1.49%	0.16%	0.05%
Pr[B|x]	0.75%	1.03%	2.45%	98.51%	99.84%	99.95%

Step 2.2: Update distribution parameters

μ_A	0.92
σ_A	0.22
p_A	49%
μ_B	3.40
σ_B	0.41
p_B	51%

Step 3.1: Calculate cluster probabilities

No change in cluster assignment → termination