

EJERCICIO DE CLÚSTER K-MEAN

Una universidad online tiene la siguiente base de datos con siete registros correspondientes a las notas de los alumnos en dos asignaturas de un máster. A partir de esta base de datos se quiere agrupar a los alumnos en dos agrupaciones exclusivas empleando el algoritmo k-means y la medida de distancia Euclídea.

	Asignatura 1	Asignatura 2
A1	1.0	1.0
A2	1.5	2.0
A3	3.0	4.0
A4	5.5	7.0
A5	3.5	5.0
A6	4.5	5.0
A7	3.5	4.5

Se sabe que en el primer paso del algoritmo k-medias se han escogido como centroides de los dos clústeres a los elementos A1 y A4. Por tanto, el valor inicial del centroides del Clúster 1 es (1.0, 2.0) y el valor inicial del Clúster 2 es (5.5, 7.0).

Aplicar el algoritmo k-medias para encontrar los elementos que componen cada una de las agrupaciones de los alumnos (Clúster 1 y Clúster 2). Describe claramente los pasos que se realizan en la ejecución del algoritmo k-medias.

Paso 2: Asignar los elementos al clúster más similar

Haciendo uso de la distancia euclidiana calculamos la distancia de los puntos a los centroides

	Punto	Distancia al centroide 1	Distancia al centroide 2	Asignación
A1	(1, 1)	0	—	Clúster 1
A2	(1.5, 2)	$\sqrt{(1.5-1)^2 + (2-1)^2} = \sqrt{1.25}$	$\sqrt{(1.5-5.5)^2 + (2-7)^2} = \sqrt{41}$	Clúster 1
A3	(3, 4)	$\sqrt{13}$	$\sqrt{15.25}$	Clúster 1
A4	(5.5, 7)	—	0	Clúster 2
A5	(3.5, 5)	$\sqrt{22.25}$	$\sqrt{8}$	Clúster 2
A6	(4.5, 5)	$\sqrt{28.25}$	$\sqrt{5}$	Clúster 2
A7	(3.5, 4.5)	$\sqrt{18.25}$	$\sqrt{10.25}$	Clúster 2

Paso 3: Recalcular las posiciones de los centroides

Clúster 1:

Punto
A1 (1, 1)
A2 (1.5, 2)
A3 (3, 4)

Clúster 2:

Punto
A4 (5.5, 7)
A5 (3.5, 5)
A6 (4.5, 5)
A7 (3.5, 4.5)

$$\text{Coord. Asig 1} = \frac{1+1.5+3}{3} = 1.833$$

$$\text{Coord. Asig 1} = \frac{5.5+3.5+4.5+3.5}{4} = 4.25$$

$$\text{Coord. Asig 2} = \frac{1+2+4}{3} = 2.333$$

$$\text{Coord. Asig 2} = \frac{7+5+5+4.5}{4} = 5.375$$

Segunda iteración

	Punto	Distancia al centroide 1	Distancia al centroide 2	Asignación
A1	(1, 1)	$\sqrt{(1-1)^2 + (1-2)^2} = \sqrt{2^2} = 2$	$\sqrt{(1-4)^2 + (1-5)^2} = \sqrt{30} \approx 5.48$	clúster 1
A2	(1, 2)	$\sqrt{1^2} = 1$	$\sqrt{(1-4)^2 + (1-5)^2} = \sqrt{30} \approx 5.48$	clúster 1
A3	(3, 4)	$\sqrt{4^2} = 4$	$\sqrt{3^2 + 4^2} = \sqrt{25} = 5$	clúster 2
A4	(5, 7)	$\sqrt{3^2 + 6^2} = \sqrt{45} \approx 6.71$	$\sqrt{4^2} = 4$	clúster 2
A5	(3, 5)	$\sqrt{2^2 + 3^2} = \sqrt{13} \approx 3.61$	$\sqrt{1^2 + 3^2} = \sqrt{10} \approx 3.16$	clúster 2
A6	(4, 5)	$\sqrt{1^2 + 2^2} = \sqrt{5} \approx 2.24$	$\sqrt{1^2 + 2^2} = \sqrt{5} \approx 2.24$	clúster 2
A7	(3, 4)	$\sqrt{2^2 + 1^2} = \sqrt{5} \approx 2.24$	$\sqrt{1^2 + 1^2} = \sqrt{2} \approx 1.41$	clúster 2

Clúster 1:

	Punto
A1	(1, 1)
A2	(1, 2)

$$\text{Coord. Asig 1} = \frac{1+1}{2} = 1$$

$$\text{Coord. Asig 2} = \frac{1+2}{2} = 1.5$$

Clúster 2:

	Punto
A3	(3, 4)
A4	(5, 7)
A5	(3, 5)
A6	(4, 5)
A7	(3, 4)

$$\text{Coord. Asig 1} = \frac{3+5+3+4+3}{5} = 4$$

$$\text{Coord. Asig 2} = \frac{4+7+5+5+4}{5} = 5.4$$

Tercera iteración. No cambia ninguno de los clústeres luego llegamos a la disposición final

	Punto	Distancia al centroide 1	Distancia al centroide 2	Asignación
A1	(1, 1)	$\sqrt{0^2 + 0^2} = 0$	$\sqrt{3^2 + 4^2} = \sqrt{25} = 5$	clúster 1
A2	(1, 2)	$\sqrt{0^2 + 1^2} = 1$	$\sqrt{2^2 + 3^2} = \sqrt{13} \approx 3.61$	clúster 1
A3	(3, 4)	$\sqrt{2^2 + 1^2} = \sqrt{5} \approx 2.24$	$\sqrt{2^2 + 2^2} = \sqrt{8} \approx 2.83$	clúster 2
A4	(5, 7)	$\sqrt{4^2 + 6^2} = \sqrt{52} \approx 7.21$	$\sqrt{5^2 + 3^2} = \sqrt{34} \approx 5.83$	clúster 2
A5	(3, 5)	$\sqrt{2^2 + 4^2} = \sqrt{20} \approx 4.47$	$\sqrt{1^2 + 2^2} = \sqrt{5} \approx 2.24$	clúster 2
A6	(4, 5)	$\sqrt{1^2 + 2^2} = \sqrt{5} \approx 2.24$	$\sqrt{0^2 + 1^2} = 1$	clúster 2
A7	(3, 4)	$\sqrt{2^2 + 1^2} = \sqrt{5} \approx 2.24$	$\sqrt{0^2 + 0^2} = 0$	clúster 2