

ALGORITMO-K-MEANS.pdf



patrivc



Apuntes Variados



4º Grado en Ingeniería Informática



Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación
Universidad de Granada



MÁSTER EN

Inteligencia Artificial & Data Management

MADRID

Formamos
talento para un futuro
Sostenible

saber más





Certifica tu nivel de inglés con LanguageCert

Language
Cert

Examen de idioma diseñado para mejorar las oportunidades académicas, profesionales y personales.

ALGORITMO K-MEANS

→ CLUSTERING

Necesita como parámetro de entrada el número de clusters deseado. Es un algoritmo iterativo; las instancias se mueven entre los clusters hasta que se alcanza el conjunto de clusters deseado.

- Ventajas:** relativamente eficiente: $O(k \cdot n)$, normalmente $k, t \ll n$. Con frecuencia finaliza en un óptimo local (dependiendo de la elección inicial de los centroides).
- Desventajas:** requiere especificar el número de clusters (k) en la entrada, sólo es aplicable cuando el concepto de media es definible, débil ante datos ruidosos y con outliers y sólo genera clusters convexos.

Ejercicio. Tenemos la siguiente BD y queremos agrupar los alumnos en dos agrupaciones estudiantiles y usando distancia Euclídea.

	Asignatura 1	Asignatura 2
A1	1	1
A2	1.5	2
A3	3	4
A4	5.5	7
A5	3.5	5
A6	4.5	5
A7	3.5	4.5

Centroides iniciales: A1 y A4

→ Cluster 1 (1, 1)
→ Cluster 2 (5.5, 7)

- Asignamos los elementos al cluster más similar. Para ello hacemos uso de la distancia Euclídea.
- Recalculamos posiciones de los centroides.
- Volvemos a iterar

1

Punto	(1, 1) distancia al Centroide 1	(5.5, 7) distancia al Centroide 2	Asignación (Asignamos al centíode más cercano)
A1 (1, 1)	0	-	Cluster 1
A2 (1.5, 2)	$\sqrt{(1.5-1)^2 + (2-1)^2} = \sqrt{1.25}$	$\sqrt{(1.5-5.5)^2 + (2-7)^2} = \sqrt{42}$	Cluster 1 (1.5, 2)
A3 (3, 4)	$\sqrt{(3-1)^2 + (4-1)^2} = \sqrt{13}$	$\sqrt{(3-5.5)^2 + (4-7)^2} = \sqrt{15.25}$	Cluster 1 (3, 4)
A4 (5.5, 7)	-	0	Cluster 2
A5 (3.5, 5)	$\sqrt{(3.5-1)^2 + (5-1)^2} = \sqrt{22.25}$	$\sqrt{(3.5-5.5)^2 + (5-7)^2} = \sqrt{8}$	Cluster 2 (4.2, 2.8)
A6 (4.5, 5)	$\sqrt{(4.5-1)^2 + (5-1)^2} = \sqrt{28.25}$	$\sqrt{(4.5-5.5)^2 + (5-7)^2} = \sqrt{5}$	Cluster 2 (3.72, 2.23)
A7 (3.5, 4.5)	$\sqrt{(3.5-1)^2 + (4.5-1)^2} = \sqrt{13.5}$	$\sqrt{(3.5-5.5)^2 + (4.5-7)^2} = \sqrt{10.25}$	Cluster 2 (4.8, 3.2)

2. Recalculamos centroides

Cluster 1

Punto	Coordenada A1 = $\frac{1+1.5+3}{3} = 1.833$
A1 (1, 1)	
A2 (1.5, 2)	
A3 (3, 4)	

Coordenada A2 = $\frac{1+2+4}{3} = 2.333$

P(1.833, 2.333)

Cluster 2

Punto	Coordenada A1 = $\frac{5.5+3.5+4.5+3.5}{4} = 4.25$
A4 (5.5, 7)	
A5 (3.5, 5)	
A6 (4.5, 5)	
A7 (3.5, 4.5)	

Coordenada A2 = $\frac{7+5+5+4.5}{4} = 5.375$

P(4.25, 5.375)



• Segunda iteración

Punto	(1', 2') Distancia Centoide 1	(4', 5') Distancia Centoide 2	Asignación
A1	$\sqrt{2^2 + 3^2}$	$\sqrt{30^2 + 25^2}$	1
A2	$\sqrt{0^2 + 8^2}$	$\sqrt{19^2 + 4^2}$	1
A3	$\sqrt{4^2 + 3^2}$	$\sqrt{3^2 + 5^2}$	2
A4	$\sqrt{35^2 + 28^2}$	$\sqrt{4^2}$	2
A5	$\sqrt{10^2 + 18^2}$	$\sqrt{0^2 + 8^2}$	2
A6	$\sqrt{14^2 + 58^2}$	$\sqrt{0^2 + 2^2}$	2
A7	$\sqrt{7^2 + 3^2}$	$\sqrt{4^2 + 5^2}$	2

Cluster 1

Punto	
A1	(1, 2)
A2	(1, 5)
C. A1 = $\frac{1+1}{2} = 1$	
C. A2 = $\frac{2+5}{3} = 2.5$	
P(1.25, 1.5)	

Cluster 2

Punto	
A3	(3, 4)
A4	(5, 2)
A5	(3, 5)
A6	(4, 5)
A7	(3, 4)
C. A1 = $\frac{3+5+3+4+3+5}{5} = 4$	
C. A2 = $\frac{4+7+5+5+4}{5} = 5.2$	
P(4.5, 2)	

• Tercera iteración

	Distancia C1.	Distancia C2.	Asignación
A1	$\sqrt{0^2 + 2^2}$	$\sqrt{25^2 + 8^2}$	1
A2	$\sqrt{0^2 + 2^2}$	$\sqrt{18^2 + 6^2}$	1
A3	$\sqrt{4^2 + 2^2}$	$\sqrt{2^2 + 2^2}$	2
A4	$\sqrt{48^2 + 2^2}$	$\sqrt{5^2 + 6^2}$	2
A5	$\sqrt{17^2 + 2^2}$	$\sqrt{0^2 + 2^2}$	2
A6	$\sqrt{22^2 + 2^2}$	$\sqrt{0^2 + 2^2}$	2
A7	$\sqrt{14^2 + 6^2}$	$\sqrt{0^2 + 2^2}$	2

Terminamos aquí porque no cambian ya los clusters