Pecotche Andres
Lucas Carballo

Minería de Datos usando Sistemas Inteligentes

PRACTICA 2 – AGRUPAMIENTO (CLUSTERING).

2.A)

	Iteración 1			Iteración 2			Iteración 3		
	c1 (1,0)	c2 (3,2)		c1 (1.5,1)	c2 (3.5,4)		c1 (1.5,1)	c2 (3.5,4)	
Puntos	Dist. a **c1**	Dist a **c2**	Centroide Asignado	Dist a **c1**	Dist a **c2**	Centroide Asignado	Dist a **c1**	Dist a **c2**	Centroide Asignado
x1 = (2,0)	1	5	C1	1.25	18.25	C1	1.25	18.25	C1
x2 = (1,2)	4	4	C1	1.25	10.25	C1	1.25	10.25	C1
x3 = (3,4)	20	4	C2	11.25	0.25	C2	11.25	0.25	C2
x4 = (4,4)	25	5	C2	15.25	0.25	C2	15.25	0.25	C2

ITERACIÓN 1:

	C1=(1,0)	C2=(3,2)
X1=(2,0)	1	5
X2=(1,2)	4	4
X3=(3,4)	20	4
X4=(4,4)	25	5

Cálculos de la primera iteración:

Calculamos la distancia euclídea al cuadrado entre los puntos y los centroides:

$d(X_N,C_M) = d((X_x,X_y),(C_x,C_y)) = (X_x-C_x)^2 - (X_y-C_y)^2$

$d(x1,c1) = d((2,0),(1,0)) = (2-1)^2+(0-0)^2 = \mathbf{1}$
$d(x1,c2) = d((2,0),(3,2)) = (2-3)^2+(0-2)^2 = 5$
$d(x2,c1) = d((1,2),(1,0)) = (1-1)^2+(2-0)^2 = \mathbf{4}$ **?**
$d(x2,c2) = d((1,2),(3,2)) = (1-3)^2+(2-2)^2 = 4$
$d(x3,c1) = d((3,4),(1,0)) = (3-1)^2+(4-0)^2 = 20$
$d(x3,c2) = d((3,4),(3,2)) = (3-3)^2+(4-2)^2 = \mathbf{4}$
$d(x4,c1) = d((4,4),(1,0)) = (4-1)^2+(4-0)^2 = 25$
$d(x4,c2) = d((4,4),(3,2)) = (4-3)^2+(4-2)^2 = \mathbf{5}$

Para calcular los nuevos centroides:

Se promedian los valores pertenecientes a cada centroide de la iteración anterior, es este caso X1 y X2 pertenecen al centroide C1, y X3 y X4 pertenecen al centroide C2.
Para esto calculamos (C_x,C_y), donde C_x sera el promedio de los XN_x y C_y será el promedio de los XN_y.

C1=(C1x,C1y)
C1x=(X1+X2)/2 = (1 + 2)/2 = **1.5**
C1y=(Y1+Y2)/2+ (0+2)/2= 2/2 =**1**
C1=(1.5,1)

C2=(C2x,C2y)
C2x=(3+4)/2= **3.5**
C2y=(4+4)/2= **4**
C2=(3.5,4)

ITERACIÓN 2:

	C1=(1.5,1)	C2=(3.5,4)
X1=(2,0)	1.25	18.25
X2=(1,2)	1.25	10.25
X3=(3,4)	11.25	0.25
X4=(4,4)	15.25	0.25

Cálculos:

d(x1,c1) = d((2,0),(1.5,1))= $(2-1.5)^2+(0-1)^2$ =**1.25**
d(x1,c2) = d((2,0),(3.5,4))= $(2-3.5)^2+(0-4)^2$ = 18.25
d(x2,c1) = d((1,2),(1.5,1))= $(1-1.5)^2+(2-1)^2$= **1.25**
d(x2,c2) = d((1,2),(3.5,4))= $(1-3.5)^2+(2-4)^2$= 10.25
d(x3,c1) = d((3,4),(1.5,1))= $(3-1.5)^2+(4-1)^2$= 11.25
d(x3,c2) = d((3,4),(3.5,4))= $(3-3.5)^2+(4-4)^2$= **0.25**
d(x4,c1) = d((4,4),(1.5,1))= $(4-1.5)^2+(4-1)^2$= 15.25
d(x4,c2) = d((4,4),(3.5,4))= $(4-3.5)^2+(4-4)^2$= **0.25**

Calculamos nuevamente los centroides:
(como nuevamente C1 está conformado por X1 y X2, y C2 por X3 y X4, los nuevos centroides son los mismos.

C1=(C1x,C1y)

$C1x=(X1_X+X2_X)/2 = (2 + 1)/2 =$ **3/2**
$C1y=(X1_Y+X2_Y)/2+ (0+2)/2= 2/2 =$**1**
C1=(1.5,1)

C2=(C2x,C2y)
$C2x=(X3_X+X4_X)/2 = (3+4)/2= 7/2$
$C2y=(X3_Y+X4_Y)/2 = (4+4)/2= 8/2=4$
C2=(3.5,4)

Como los centroides resultantes de la iteración 2 son los mismos que los de la iteración 1, de ahora en adelante los resultados seguirán siendo los mismos, y ya no es necesario continuar iterando.

	C1=(1.5,1)	C2=(3.5,4)
X1=(2,0)	1.25	18.25
X2=(1,2)	1.25	10.25
X3=(3,4)	11.25	0.25
X4=(4,4)	15.25	0.25

C1=(C1x,C1y)
C1x=(X1+X2)/2 = (2 + 1)/2 = **3/2**
C1y=(Y1+Y2)/2+ (0+2)/2= 2/2 =**1**
C1=(1.5,1)

C2=(C2x,C2y)
C2x=(3+4)/2= 7/2
C2y=(4+4)/2= 8/2=4
C2=(3.5,4)

B) Como los centroides asignados a los puntos, no varían, los centroides resultantes van a seguir siendo los mismos y por tanto seguirán permaneciendo los mismos grupos asignados, siendo un bucle en infinito en el que no hay cambios. Por tanto, en cuanto el resultado de una iteración provee los mismos centroides que la iteración anterior, se dice que la asignación de los grupos ya convergio.

C)

	Iteracion 1			Iteracion 2		
	C1=(1,2)	C2=(3,4)		C1=(1.5,1)	C2=(3.5,4)	
Puntos	Dist. a C1	Dist. a C2	Cent. asignado	Dist. a C1	Dist. a C2	Cent. asignado
X1=(2,0)	5	17	C1	1.25	18.25	C1
X2=(1,2)	0	8	C1	1.25	10.25	C1
X3=(3,4)	8	0	C2	11.25	0.25	C2
X4=(4,4)	13	1	C2	15.25	0.25	C2

En este caso se requieren dos iteraciones debido a que cuando calculamos la segunda iteración con los nuevos centroides, los grupos conformados siguen siendo los mismos (x1 y x2 en C1, y X2 y X3 en C2) por lo que el cálculo de los centroides para una siguiente iteración daría como resultado los mismos centroides que en la iteración 2, haciendo que la conformación de los grupos no cambie.

El resultado final, es el mismo que el del inciso A).

Cálculos:

Iteración 1:

$d(x1,c1) = d((2,0),(1,2)) = (2-1)^2+(0-2)^2 = \mathbf{5}$
$d(x1,c2) = d((2,0),(3,4)) = (2-3)^2+(0-4)^2 = 17$
$d(x2,c1) = d((1,2),(1,2)) = (1-1)^2+(2-2)^2 = \mathbf{0}$
$d(x2,c2) = d((1,2),(3,4)) = (1-3)^2+(2-4)^2 = 8$
$d(x3,c1) = d((3,4),(1,2)) = (3-1)^2+(4-2)^2 = 8$
$d(x3,c2) = d((3,4),(3,4)) = (3-3)^2+(4-4)^2 = \mathbf{0}$
$d(x4,c1) = d((4,4),(1,2)) = (4-1)^2+(4-2)^2 = 13$
$d(x4,c2) = d((4,4),(3,4)) = (4-3)^2+(4-4)^2 = \mathbf{1}$

Centroides resultantes:

C1=(C1x,C1y)
$C1x=(X1_X+X2_X)/2 = (2 + 1)/2 = \mathbf{1.5}$
$C1y=(X1_Y+X2_Y)/2+ (0+2)/2= 2/2 =\mathbf{1}$
C1=(1.5,1)

C2=(C2x,C2y)
$C2x=(X3_X+X4_X)/2 = (3+4)/2= \mathbf{3.5}$
$C2y=(X3_Y+X4_Y)/2 = (4+4)/2= \mathbf{4}$
C2=(3.5,4)

Iteración 2:

$d(x1,c1) = d((2,0),(1.5,1))= (2-1.5)^2+(0-1)^2 = \mathbf{1.25}$
$d(x1,c2) = d((2,0),(3.5,4))= (2-3.5)^2+(0-4)^2 = 18.25$
$d(x2,c1) = d((1,2),(1.5,1))= (1-1.5)^2+(2-1)^2= \mathbf{1.25}$
$d(x2,c2) = d((1,2),(3.5,4))= (1-3.5)^2+(2-4)^2= 10.25$
$d(x3,c1) = d((3,4),(1.5,1))= (3-1.5)^2+(4-1)^2= 11.25$
$d(x3,c2) = d((3,4),(3.5,4))= (3-3.5)^2+(4-4)^2= \mathbf{0.25}$
$d(x4,c1) = d((4,4),(1.5,1))= (4-1.5)^2+(4-1)^2= 15.25$
$d(x4,c2) = d((4,4),(3.5,4))= (4-3.5)^2+(4-4)^2= \mathbf{0.25}$

D)

	Iteracion 1			Iteracion 2			Iteracion 3		
	C1=(1,4)	C2=(5,6)		C1=(2,2)	C2=(4,4)		C1=(1.5,1)	C2=(3.5,4)	
Puntos	Dist. a C1	Dist. a C2	Cent. asignado	Dist. a C1	Dist. a C2	Cent. asignado	Dist. a C1	Dist. a C2	Cent. asignado
X1=(2,0)	17	45	C1	4	20	C1	1.25	18.25	C1
X2=(1,2)	4	32	C1	1	13	C1	1.25	10.25	C1
X3=(3,4)	4	8	C1	5	1	C2	11.25	0.25	C2
X4=(4,4)	9	5	C2	8	0	C2	15.25	0.25	C2

Nuevamente volvimos a obtener el mismo resultado final, solo que esta vez se requirió una tercera iteración hasta que los resultados de los grupos convergieron.

Cálculos:
Iteración 1:

$d(x1,c1) = d((2,0),(1,4))= (2-1)^2+(0-4)^2 = \mathbf{17}$
$d(x1,c2) = d((2,0),(5,6))= (2-5)^2+(0-6)^2 = 45$
$d(x2,c1) = d((1,2),(1,4))= (1-1)^2+(2-4)^2= \mathbf{4}$
$d(x2,c2) = d((1,2),(5,6))= (1-5)^2+(2-6)^2= 32$
$d(x3,c1) = d((3,4),(1,4))= (3-1)^2+(4-5)^2= \mathbf{4}$
$d(x3,c2) = d((3,4),(5,6))= (3-5)^2+(4-6)^2= 8$
$d(x4,c1) = d((4,4),(1,4))= (4-1)^2+(4-4)^2= 9$
$d(x4,c2) = d((4,4),(5,6))= (4-5)^2+(4-6)^2= \mathbf{5}$

Centroides resultantes:

$C1=(C1x,C1y)$
$C1x=(X1_X+X2_X+X3_X)/3 = (2 + 1 + 3)/3 = \mathbf{2}$
$C1y=(X1_Y+X2_Y+X3_Y)/3+ (0+2+4)/3 =\mathbf{2}$
C1=(2,2)

$C2=(C2x,C2y)$
$C2x=(X4_X)/2 = \mathbf{4}$
$C2y=(X4_Y)/2 = \mathbf{4}$

C2=(4,4)

Iteración 2:

d(x1,c1) = d((2,0),(2,2))= $(2-2)^2+(0-2)^2$ = **4**
d(x1,c2) = d((2,0),(4,4))= $(2-4)^2+(0-4)^2$ = 20
d(x2,c1) = d((1,2),(2,2))= $(1-2)^2+(2-2)^2$= **1**
d(x2,c2) = d((1,2),(4,4))= $(1-4)^2+(2-4)^2$= 13
d(x3,c1) = d((3,4),(2,2))= $(3-2)^2+(4-2)^2$= 5
d(x3,c2) = d((3,4),(4,4))= $(3-4)^2+(4-4)^2$= **1**
d(x4,c1) = d((4,4),(2,2))= $(4-2)^2+(4-2)^2$= 8
d(x4,c2) = d((4,4),(4,4))= $(4-4)^2+(4-4)^2$= **0**

Centroides resultantes:

C1=(C1x,C1y)
C1x=($X1_X$+$X2_X$)/2 = (2 + 1)/2 = **1.5**
C1y=($X1_Y$+$X2_Y$)/2+ (0+2)/2= 2/2 =**1**
C1=(1.5,1)

C2=(C2x,C2y)
C2x=($X3_X$+$X4_X$)/2 = (3+4)/2= **3.5**
C2y=($X3_Y$+$X4_Y$)/2 = (4+4)/2= **4**
C2=(3.5,4)

Iteración 3:

d(x1,c1) = d((2,0),(1.5,1))= $(2-1.5)^2+(0-1)^2$ =**1.25**
d(x1,c2) = d((2,0),(3.5,4))= $(2-3.5)^2+(0-4)^2$ = 18.25
d(x2,c1) = d((1,2),(1.5,1))= $(1-1.5)^2+(2-1)^2$= **1.25**
d(x2,c2) = d((1,2),(3.5,4))= $(1-3.5)^2+(2-4)^2$= 10.25
d(x3,c1) = d((3,4),(1.5,1))= $(3-1.5)^2+(4-1)^2$= 11.25
d(x3,c2) = d((3,4),(3.5,4))= $(3-3.5)^2+(4-4)^2$= **0.25**
d(x4,c1) = d((4,4),(1.5,1))= $(4-1.5)^2+(4-1)^2$= 15.25
d(x4,c2) = d((4,4),(3.5,4))= $(4-3.5)^2+(4-4)^2$= **0.25**

Los grupos asignados vuelven a ser los mismos que la iteración anterior por lo que no se realizan más iteraciones.

E)
Indique cuáles de estas cosas pueden variar si se corre el algoritmo K-Medias con K=4 en un conjunto de datos fijo y se varían los valores de los centroides iniciales:
Cantidad de iteraciones - Valores de los centroides finales - Cantidad de centroides - Cantidad de grupos encontrados - Cantidad de ejemplos por centroide/grupo.

Cantidad de iteraciones: La cantidad de iteraciones, puede variar dependiendo de cómo se posicionan los valores de los centroides iniciales. Por ejemplo, en el ejercicio del inciso A, si los valores iniciales de los centroides fueran C1=(1.5,1) y C2=(3.5,4), en una iteración se resolvería el agrupamiento. En cambio si los centros son como los del inciso D, se puede observar que se requirió una iteración más para llegar hasta la convergencia.

Valores de los centroides finales: Los valores de los centroides finales (y por tanto los grupos finales), pueden variar dependiendo de la ubicación de los centros iniciales.

Cantidad de centroides: La cantidad de centroides no puede variar, ya que si queremos obtener 4 grupos (k=4) debemos utilizar 4 centroides.

Cantidad de grupos: Esto no puede cambiar ya que está especificado en el enunciado que es un valor fijo (k=4).

Cantidad de ejemplos por centroide/grupo: Al igual que los valores de los centroides finales, la cantidad de elementos por centroide, también puede variar, justamente debido a que el primero puede, y si los valores de los centroides finales cambia, junto con estos, pueden cambiar el número de ejemplos por grupo.