

Cuaderno de trabajo

Clustering: algoritmo C-medias

Albert Sanchis

Departamento de Sistemas Informáticos y Computación

Objetivos formativos

■ Aplicar el algoritmo *C*-medias de Duda y Hart



Algoritmo C-medias de Duda y Hart [1]

```
Algorithm C-means
Input: X; C; \Pi = \{X_1, \dots, X_C\};
Output: \Pi^* = \{X_1, \dots, X_C\}; \, m_1, \dots, m_C; \, J
for c=1 to C do {m m}_c=rac{1}{n_c}\sum_{{m x}\in X_c}{m x} endfor
repeat
   transfers = false
   forall x \in X (let i : x \in X_i) do
      if n_i > 1 then
        j^* = \operatorname*{arg\,min}_{i \neq i} \frac{n_j}{n_i + 1} \| oldsymbol{x} - oldsymbol{m}_j \|^2
        \Delta J = \frac{n_{j^*}}{n_{i^*} + 1} \| \boldsymbol{x} - \boldsymbol{m}_{j^*} \|^2 - \frac{n_i}{n_i - 1} \| \boldsymbol{x} - \boldsymbol{m}_i \|^2
         if \triangle J < 0 then
           transfers = true
           oldsymbol{m}_i = oldsymbol{m}_i - rac{oldsymbol{x} - oldsymbol{m}_i}{n_i - 1} \quad oldsymbol{m}_{j^*} = oldsymbol{m}_{j^*} + rac{oldsymbol{x} - oldsymbol{m}_{j^*}}{n_{i^*} + 1}
           X_i = X_i - \{x\} X_{i^*} = X_{i^*} + \{x\}
           J = J + \wedge J
         endif
      endif
   endforall
until \neg transfers
```



Algoritmo C-medias de Duda y Hart

- *Entrada:* una partición inicial, $\Pi = \{X_1, \dots, X_C\}$
- *Salida:* una partición optimizada, $\Pi^* = \{X_1, \dots, X_C\}$
- Método:

Calcular medias y J

repetir

para todo dato x

Sea i el clúster en el que se encuentra $oldsymbol{x}$

Hallar un $j^* \neq i$ que minimice $\triangle J$ al transferir \boldsymbol{x} de i a j^*

Si $\triangle J < 0$, transferir \boldsymbol{x} de i a j^* y actualizar medias y J

hasta no encontrar transferencias provechosas

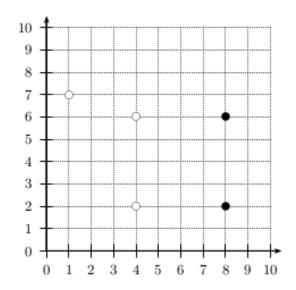


■ Cuestión 1: Dados los siguientes 5 vectores bidimensionales:

$$\boldsymbol{x}_1 = \begin{pmatrix} 1 \\ 7 \end{pmatrix}$$
 $\boldsymbol{x}_2 = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$ $\boldsymbol{x}_3 = \begin{pmatrix} 4 \\ 6 \end{pmatrix}$ $\boldsymbol{x}_4 = \begin{pmatrix} 8 \\ 2 \end{pmatrix}$ $\boldsymbol{x}_5 = \begin{pmatrix} 8 \\ 6 \end{pmatrix}$

y la siguiente partición inicial en dos clústers:

$$\Pi = \{X_1 = \{x_1, x_2, x_3\}, X_2 = \{x_4, x_5\}\}$$



¿Cuál es la partición Π^* resultante tras aplicar el algoritmo C-medias de Duda y Hart?



```
Algorithm C-means
Input: X; C; \Pi = \{X_1, \dots, X_C\};
Output: \Pi^* = \{X_1, \dots, X_C\}; \, m_1, \dots, m_C; \, J
for c=1 to C do {m m}_c=\frac{1}{n_c}\sum_{{m x}\in X_c}{m x} endfor
repeat
   transfers = false
   forall x \in X (let i : x \in X_i) do
      if n_i > 1 then
         j^* = \operatorname*{arg\,min}_{i \neq i} \frac{n_j}{n_i + 1} \| oldsymbol{x} - oldsymbol{m}_j \|^2
         \Delta J = \frac{n_{j^*}}{n_{i^*} + 1} \| \boldsymbol{x} - \boldsymbol{m}_{j^*} \|^2 - \frac{n_i}{n_i - 1} \| \boldsymbol{x} - \boldsymbol{m}_i \|^2
         if \triangle J < 0 then
            transfers = true
            egin{aligned} m{m}_i &= m{m}_i - rac{m{x} - m{m}_i}{n_i - 1} & m{m}_{j^*} &= m{m}_{j^*} + rac{m{x} - m{m}_{j^*}}{n_{j^*} + 1} \ X_i &= X_i - \{m{x}\} & X_{j^*} &= X_{j^*} + \{m{x}\} \end{aligned}
             J = J + \triangle J
         endif
      endif
   endforall
until \neg transfers
```

П	$\{X_1 = \{x_1, x_2, x_3\}, X_2 = \{x_4, x_5\}\}$
m_1	
m_2	
J_1	
J_2	
J	
زTra	insferimos $oldsymbol{x}_1=(1,7)^t$ de X_1 a X_2 ?
j^*	
$\triangle J$	
	insferimos $\boldsymbol{x}_2=(4,2)^t$ de X_1 a X_2 ?
j^*	
$\triangle J$	
m_1	
m_2	
Π	
J	
	insferimos $\boldsymbol{x}_3=(4,6)^t$ de X_1 a X_2 ?
j^*	
$\triangle J$	
	insferimos $\boldsymbol{x}_4 = (8,2)^t$ de X_2 a X_1 ?
j^*	
$\triangle J$	
	Insferimos $\boldsymbol{x}_5 = (8,6)^t$ de X_2 a X_1 ?
j^*	
$\triangle J$	



Input:
$$X; C; \Pi = \{X_1, ..., X_C\};$$

Output:
$$\Pi^* = \{X_1, \dots, X_C\}; \, {\bm m}_1, \dots, {\bm m}_C; \, J$$

for
$$c=1$$
 to C do ${\boldsymbol m}_c=\frac{1}{n_c}\sum_{{\boldsymbol x}\in X_c}{\boldsymbol x}$ endfor repeat

$$transfers = false$$

forall
$$x \in X$$
 (let $i : x \in X_i$) do

if
$$n_i > 1$$
 then

$$j^* = \operatorname*{arg\,min}_{j
eq i} \frac{n_j}{n_i + 1} \| \boldsymbol{x} - \boldsymbol{m}_j \|^2$$

$$\triangle J = \frac{n_{j^*}}{n_{i^*} + 1} \| \boldsymbol{x} - \boldsymbol{m}_{j^*} \|^2 - \frac{n_i}{n_i - 1} \| \boldsymbol{x} - \boldsymbol{m}_i \|^2$$

if $\triangle J < 0$ then

$$transfers = true$$

$$egin{aligned} m{m}_i &= m{m}_i - rac{m{x} - m{m}_i}{n_i - 1} & m{m}_{j^*} &= m{m}_{j^*} + rac{m{x} - m{m}_{j^*}}{n_{j^*} + 1} \ X_i &= X_i - \{m{x}\} & X_{j^*} &= X_{j^*} + \{m{x}\} \end{aligned}$$

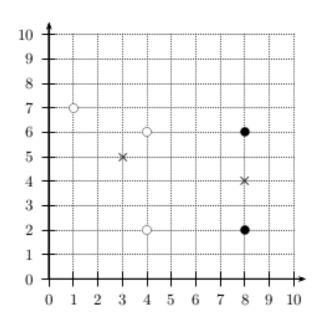
$$J = J + \triangle J$$

endif

endif

endforall

П	$\{X_1 = \{m{x}_1, m{x}_2, m{x}_3\}, X_2 = \{m{x}_4, m{x}_5\}\}$
m_1	$\begin{pmatrix} 3 \\ 5 \end{pmatrix}$
m_2	$\begin{pmatrix} 8 \\ 4 \end{pmatrix}$
J_1	20
J_2	8
J	28
	insferimos $\boldsymbol{x}_1=(1,7)^t$ de X_1 a X_2 ?
j^*	2
$\triangle J$	80/3





Input:
$$X$$
; C ; $\Pi = \{X_1, \dots, X_C\}$;

Output:
$$\Pi^* = \{X_1, \dots, X_C\}; \, {\bm m}_1, \dots, {\bm m}_C; \, J$$

for
$$c=1$$
 to C do ${m m}_c=\frac{1}{n_c}\sum_{{m x}\in X_c}{m x}$ endfor repeat

transfers = false

forall
$$x \in X$$
 (let $i : x \in X_i$) do

if $n_i > 1$ then

$$j^* = \operatorname*{arg\,min}_{j
eq i} \, rac{n_j}{n_j + 1} \, \|oldsymbol{x} - oldsymbol{m}_j\|^2$$

$$\Delta J = \frac{n_{j^*}}{n_{j^*} + 1} \| \boldsymbol{x} - \boldsymbol{m}_{j^*} \|^2 - \frac{n_i}{n_i - 1} \| \boldsymbol{x} - \boldsymbol{m}_i \|^2$$

if $\triangle J < 0$ then

$$transfers = true$$

$$egin{aligned} m{m}_i &= m{m}_i - rac{m{x} - m{m}_i}{n_i - 1} & m{m}_{j^*} &= m{m}_{j^*} + rac{m{x} - m{m}_{j^*}}{n_{j^*} + 1} \ X_i &= X_i - \{m{x}\} & X_{j^*} &= X_{j^*} + \{m{x}\} \end{aligned}$$

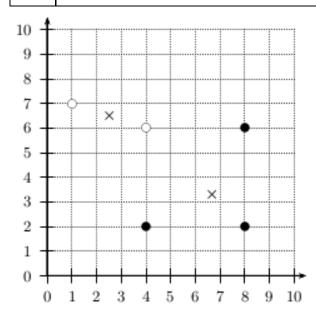
$$J = J + \triangle J$$

endif

endif

endforall

¿Transferimos $x_2 = (4, 2)^t$ de X_1 a X_2 ?	
j^*	2
$\triangle J$	-5/3
m_1	$(2,5,6,5)^t$
m_2	$(6,67,3,33)^t$
П	$\{X_1 = \{x_1, x_3\}, X_2 = \{x_2, x_4, x_5\}\}$
J	26,33
¿Tra	insferimos $\boldsymbol{x}_3=(4,6)^t$ de X_1 a X_2 ?
j^*	2
$\triangle J$	5,67
_	insferimos $\boldsymbol{x}_4 = (8,2)^t$ de X_2 a X_1 ?
j^*	1
$\triangle J$	28,33
¿Tra	insferimos $oldsymbol{x}_5 = (8,6)^t$ de X_2 a X_1 ?
j^*	1
$\triangle J$	7



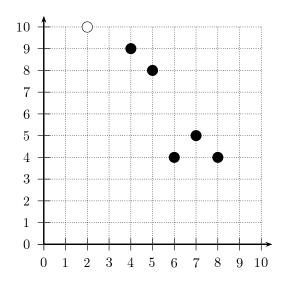


■ Cuestión 2: Dados los siguientes 6 vectores bidimensionales:

$$\mathbf{x}_1 = \begin{pmatrix} 2 \\ 10 \end{pmatrix} \mathbf{x}_2 = \begin{pmatrix} 8 \\ 4 \end{pmatrix} \mathbf{x}_3 = \begin{pmatrix} 5 \\ 8 \end{pmatrix} \mathbf{x}_4 = \begin{pmatrix} 7 \\ 5 \end{pmatrix} \mathbf{x}_5 = \begin{pmatrix} 6 \\ 4 \end{pmatrix} \mathbf{x}_6 = \begin{pmatrix} 4 \\ 9 \end{pmatrix}$$

y la siguiente partición inicial en dos clústers:

$$\Pi = \{X_1 = \{x_1\}, X_2 = \{x_2, x_3, x_4, x_5, x_6\}\}$$



¿Cuál es la partición Π^* resultante tras aplicar el algoritmo C-medias de Duda y Hart?

Input:
$$X; C; \Pi = \{X_1, \dots, X_C\};$$

Output:
$$\Pi^* = \{X_1, \dots, X_C\}; \, m_1, \dots, m_C; \, J$$

for
$$c=1$$
 to C do ${m m}_c=\frac{1}{n_c}\sum_{{m x}\in X_c}{m x}$ endfor repeat

$$transfers = false$$

forall
$$x \in X$$
 (let $i : x \in X_i$) do

if $n_i > 1$ then

$$j^* = rg \min_{j
eq i} rac{n_j}{n_j + 1} \|oldsymbol{x} - oldsymbol{m}_j\|^2$$

$$\Delta J = \frac{n_{j^*}}{n_{j^*} + 1} \| \boldsymbol{x} - \boldsymbol{m}_{j^*} \|^2 - \frac{n_i}{n_i - 1} \| \boldsymbol{x} - \boldsymbol{m}_i \|^2$$

if $\triangle J < 0$ then

$$transfers = true$$

$$egin{aligned} m{m}_i &= m{m}_i - rac{m{x} - m{m}_i}{n_i - 1} & m{m}_{j^*} &= m{m}_{j^*} + rac{m{x} - m{m}_{j^*}}{n_{j^*} + 1} \ X_i &= X_i - \{m{x}\} & X_{j^*} &= X_{j^*} + \{m{x}\} \end{aligned}$$

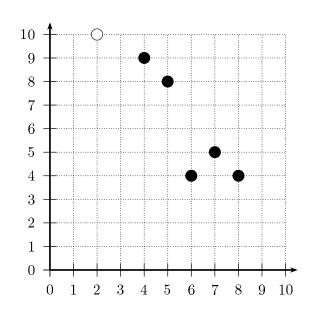
$$J = J + \triangle J$$

endif

endif

endforall

Π	$\{X_1 = \{x_1\}, X_2 = \{x_2, x_3 x_4, x_5, x_6\}\}$
m_1	
m_2	
J_1	
J_2	
J	
Τς	ransferimos $oldsymbol{x}_2=(8,4)^t$ de X_2 a X_1 ?
$\triangle J$	
Τς	ransferimos $oldsymbol{x}_3=(5,8)^t$ de X_2 a X_1 ?
$\triangle J$	
Τς	ransferimos $oldsymbol{x}_4=(7,5)^t$ de X_2 a X_1 ?
$\triangle J$	
T	ransferimos $oldsymbol{x}_5=(6,4)^t$ de X_2 a X_1 ?
$\triangle J$	
Τς	ransferimos $oldsymbol{x}_6=(4,9)^t$ de X_2 a X_1 ?
$\triangle J$	





Input:
$$X; C; \Pi = \{X_1, \dots, X_C\};$$

Output:
$$\Pi^* = \{X_1, \dots, X_C\}; \, m_1, \dots, m_C; \, J$$

for
$$c=1$$
 to C do ${m m}_c=\frac{1}{n_c}\sum_{{m x}\in X_c}{m x}$ endfor repeat

transfers = false

forall
$$x \in X$$
 (let $i : x \in X_i$) do

if $n_i > 1$ then

$$j^* = \operatorname*{arg\,min}_{j
eq i} \, rac{n_j}{n_j + 1} \, \lVert oldsymbol{x} - oldsymbol{m}_j
Vert^2$$

$$\triangle J = rac{n_{j^*}}{n_{i^*} + 1} \|m{x} - m{m}_{j^*}\|^2 - rac{n_i}{n_i - 1} \|m{x} - m{m}_i\|^2$$

if $\triangle J < 0$ then

$$transfers = true$$

$$egin{aligned} oldsymbol{m}_i &= oldsymbol{m}_i - rac{oldsymbol{x} - oldsymbol{m}_i}{n_i - 1} & oldsymbol{m}_{j^*} &= oldsymbol{m}_{j^*} + rac{oldsymbol{x} - oldsymbol{m}_{j^*}}{n_{j^*} + 1} \ X_i &= X_i - \{oldsymbol{x}\} & X_{j^*} &= X_{j^*} + \{oldsymbol{x}\} \end{aligned}$$

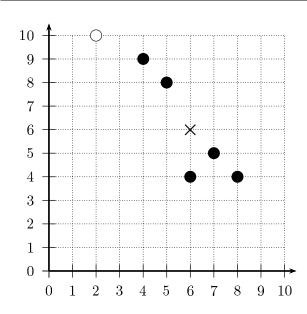
$$J = J + \triangle J$$

endif

endif

endforall

Π	$\{X_1 = \{x_1\}, X_2 = \{x_2, x_3 x_4, x_5, x_6\}\}$
m_1	$(2,10)^t$
m_2	$(6,6)^t$
J_1	0
J_2	32
J	32
Τς	ransferimos $oldsymbol{x}_2=(8,4)^t$ de X_2 a X_1 ?
$\triangle J$	26
Τς	ransferimos $oldsymbol{x}_3=(5,8)^t$ de X_2 a X_1 ?
$\triangle J$	0.25
Τς	ransferimos $oldsymbol{x}_4=(7,5)^t$ de X_2 a X_1 ?
$\triangle J$	47.5
Τς	ransferimos $oldsymbol{x}_5=(6,4)^t$ de X_2 a X_1 ?
$\triangle J$	21
Τς	ransferimos $oldsymbol{x}_6=(4,9)^t$ de X_2 a X_1 ?
$\triangle J$	-13.75





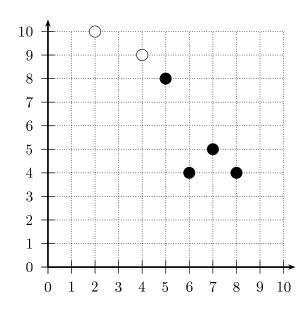
Input:
$$X$$
; C ; $\Pi = \{X_1, \dots, X_C\}$; Output: $\Pi^* = \{X_1, \dots, X_C\}$; m_1, \dots, m_C ; J for $c=1$ to C do $m_c = \frac{1}{n_c} \sum_{x \in X_c} x$ endfor repeat
$$transfers = \text{false}$$
 forall $x \in X$ (let $i: x \in X_i$) do if $n_i > 1$ then
$$j^* = \arg\min_{j \neq i} \frac{n_j}{n_j + 1} \|x - m_j\|^2$$

$$\triangle J = \frac{n_{j^*}}{n_{j^*} + 1} \|x - m_{j^*}\|^2 - \frac{n_i}{n_i - 1} \|x - m_i\|^2$$
 if $\triangle J < 0$ then
$$transfers = \text{true}$$

$$m_i = m_i - \frac{x - m_i}{n_i - 1} \qquad m_{j^*} = m_{j^*} + \frac{x - m_{j^*}}{n_{j^*} + 1}$$

$$X_i = X_i - \{x\} \qquad X_{j^*} = X_{j^*} + \{x\}$$
 endifendifendiforall

Π	$\{X_1 = \{x_1, x_6\}, X_2 = \{x_2, x_3 x_4, x_5\}\}$
m_1	
m_2	
J	
Trخ	ansferimos $\boldsymbol{x}_1=(2,10)^t$ de X_1 a X_2 ?
$\triangle J$	
Ti	ransferimos $oldsymbol{x}_2=(8,4)^t$ de X_2 a X_1 ?
$\triangle J$	
Τ̈́ς	ransferimos $oldsymbol{x}_3=(5,8)^t$ de X_2 a X_1 ?
$\triangle J$	





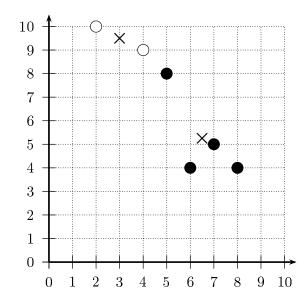
Input:
$$X; C; \Pi = \{X_1, \dots, X_C\};$$
Output: $\Pi^* = \{X_1, \dots, X_C\}; m_1, \dots, m_C; J$
for $c = 1$ to C do $m_c = \frac{1}{n_c} \sum_{x \in X_c} x$ endfor repeat $transfers = \text{false}$ forall $x \in X$ (let $i : x \in X_i$) do

if $n_i > 1$ then
$$j^* = \arg\min_{j \neq i} \frac{n_j}{n_j + 1} \|x - m_j\|^2$$

$$\triangle J = \frac{n_{j^*}}{n_{j^*} + 1} \|x - m_{j^*}\|^2 - \frac{n_i}{n_i - 1} \|x - m_i\|^2$$
 if $\triangle J < 0$ then $transfers = \text{true}$
$$m_i = m_i - \frac{x - m_i}{n_i - 1} \qquad m_{j^*} = m_{j^*} + \frac{x - m_{j^*}}{n_{j^*} + 1}$$

$$X_i = X_i - \{x\} \qquad X_{j^*} = X_{j^*} + \{x\}$$
 endifendifendiforall

П	$\{X_1 = \{x_1, x_6\}, X_2 = \{x_2, x_3 x_4, x_5\}\}$
m_1	$(3, 9, 5)^t$
m_2	$(6,5,5,25)^t$
J	18,25
¿Transferimos $oldsymbol{x}_1=(2,10)^t$ de X_1 a X_2 ?	
$\triangle J$	31.75
Τς	ransferimos $\boldsymbol{x}_2 = (8,4)^t$ de X_2 a X_1 ?
$\triangle J$	31.75
$oldsymbol{\mathcal{L}}$ Transferimos $oldsymbol{x}_3=(5,8)^t$ de X_2 a X_1 ?	
$\triangle J$	-8.91





Input:
$$X$$
; C ; $\Pi = \{X_1, \dots, X_C\}$;
Output: $\Pi^* = \{X_1, \dots, X_C\}$; m_1, \dots, m_C ; J for $c=1$ to C do $m_c = \frac{1}{n_c} \sum_{x \in X_c} x$ endfor repeat
$$transfers = \text{false}$$
 forall $x \in X$ (let $i: x \in X_i$) do
$$\text{if } n_i > 1 \text{ then}$$

$$j^* = \arg\min_{j \neq i} \frac{n_j}{n_j + 1} \|x - m_j\|^2$$

$$\triangle J = \frac{n_{j^*}}{n_{j^*} + 1} \|x - m_{j^*}\|^2 - \frac{n_i}{n_i - 1} \|x - m_i\|^2$$

$$\text{if } \triangle J < 0 \text{ then}$$

$$transfers = \text{true}$$

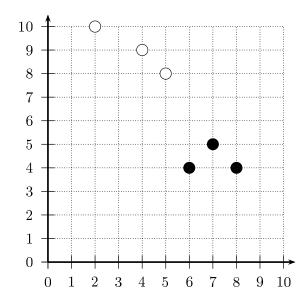
$$m_i = m_i - \frac{x - m_i}{n_i - 1}$$

$$m_{j^*} = m_{j^*} + \frac{x - m_{j^*}}{n_{j^*} + 1}$$

$$X_i = X_i - \{x\}$$

$$J = J + \triangle J$$
 endiferendification.

Π	$\{X_1 = \{x_1, x_3, x_6\}, X_2 = \{x_2, x_4, x_5\}\}$
m_1	
m_2	
J	
Tخ	ransferimos $\boldsymbol{x}_4 = (7,5)^t$ de X_2 a X_1 ?
$\triangle J$	
T	ransferimos $oldsymbol{x}_5 = (6,4)^t$ de X_2 a X_1 ?
$\triangle J$	
T	ransferimos $\boldsymbol{x}_6 = (4,9)^t$ de X_1 a X_2 ?
$\triangle J$	





Input:
$$X$$
; C ; $\Pi = \{X_1, \dots, X_C\}$;

Output: $\Pi^* = \{X_1, \dots, X_C\}$; m_1, \dots, m_C ; J

for $c = 1$ to C do $m_c = \frac{1}{n_c} \sum_{x \in X_c} x$ endfor repeat $transfers = \text{false}$ forall $x \in X$ (let $i : x \in X_i$) do

if $n_i > 1$ then
$$j^* = \arg\min_{j \neq i} \frac{n_j}{n_j + 1} \|x - m_j\|^2$$

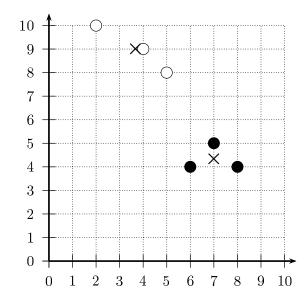
$$\triangle J = \frac{n_{j^*}}{n_{j^*} + 1} \|x - m_{j^*}\|^2 - \frac{n_i}{n_i - 1} \|x - m_i\|^2$$
 if $\triangle J < 0$ then $transfers = \text{true}$
$$m_i = m_i - \frac{x - m_i}{n_i - 1}$$

$$m_{j^*} = m_{j^*} + \frac{x - m_{j^*}}{n_{j^*} + 1}$$

$$X_i = X_i - \{x\}$$

$$J = J + \triangle J$$
 endifeendifeendforall

П	$\{X_1 = \{x_1, x_3, x_6\}, X_2 = \{x_2, x_4, x_5\}\}$	
m_1	$(3,67,9)^t$	
m_2	$(7,4,33)^t$	
J	8,67	
¿Transferimos ${m x}_4=(7,5)^t$ de X_2 a X_1 ?		
$\triangle J$	20.32	
Τς	¿Transferimos $oldsymbol{x}_5=(6,4)^t$ de X_2 a X_1 ?	
$\triangle J$	21.16	
ن Transferimos $oldsymbol{x}_6=(4,9)^t$ de X_1 a X_2 ?		
$\triangle J$	22.94	





Referencias

[1] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis*. Wiley, 1973.

