



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Quadern de treball

Clustering: algorisme C-mitjanes

Albert Sanchis

DSIC

Departament de Sistemes
Informàtics i Computació

Objectius formatius

- Aplicar l'algorisme C -mitjanes de Duda i Hart

Algorisme *C*-mitjanes de Duda i Hart [1]

Algorithm *C*-means

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

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

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

repeat

$transfers = \text{false}$

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

if $n_i > 1$ **then**

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

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

if $\Delta J < 0$ **then**

$transfers = \text{true}$

$$\mathbf{m}_i = \mathbf{m}_i - \frac{x - \mathbf{m}_i}{n_i - 1} \quad \mathbf{m}_{j^*} = \mathbf{m}_{j^*} + \frac{x - \mathbf{m}_{j^*}}{n_{j^*} + 1}$$

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

$$J = J + \Delta J$$

endif

endif

endforall

until $\neg transfers$

Algorisme *C*-mitjanes de Duda i Hart

- **Entrada:** una partició inicial, $\Pi = \{X_1, \dots, X_C\}$
- **Eixida:** una partició optimitzada, $\Pi^* = \{X_1, \dots, X_C\}$

- **Mètode:**

Calcular mitjanes i J

repetir

per a tota dada x

Siga i el clúster en el qual es troba x

Trobar un $j^* \neq i$ que minimitze ΔJ en transferir x d' i a j^*

Si $\Delta J < 0$, transferir x d' i a j^* i actualitzar mitjanes i J

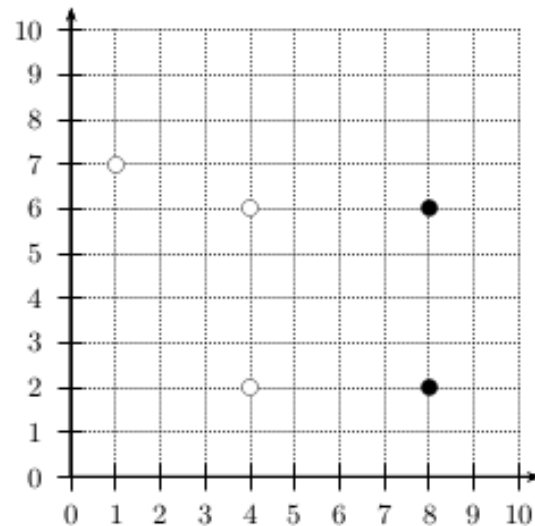
fins a no trobar transferències profitoses

- **Qüestió 1:** Donats els següents 5 vectors bidimensionals:

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

i la següent partició inicial en dos clústers:

$$\Pi = \{X_1 = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}, X_2 = \{\mathbf{x}_4, \mathbf{x}_5\}\}$$



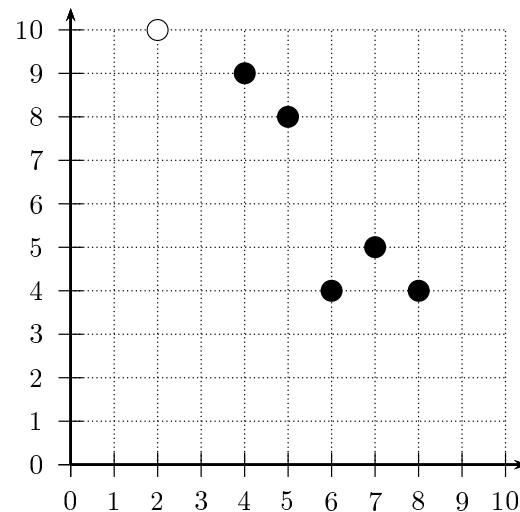
Quina és la partició Π^* resultant després d'aplicar l'algorisme *C*-mitjanes de Duda i Hart?

- **Qüestió 2:** Donats els següents 6 vectors bidimensionals:

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

i la següent partició inicial en dos clústers:

$$\Pi = \{X_1 = \{\mathbf{x}_1\}, X_2 = \{\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_6\}\}$$



Quina és la partició Π^* resultant després d'aplicar l'algorisme *C*-mitjanes de Duda i Hart?

Referències

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