3.2. El método del «vecino más cercano»

- ✔ Disponemos de una tabla resumen de tipo T(n,p)
- ✔ Los elementos de T(n,p) presentan una estructura de grupo o de jerarquía de grupos encajados.

Aplicamos las etapas ya vistas del proceso de clasificación :

Primera etapa :

Con una distancia d_{ij} podemos evaluar la disimilaridad entre los objetos a clasificar.

Podemos crear una tabla D(n,n), simétrica, que resume las distancias entre los n objetos a clasificar, comparados dos a dos.

Suponemos que es aceptable considerar que la distancia entre dos clases que contienen un solo objeto cada una es igual a la distancia entre los objetos:

$$\mathbf{d}_{(\{x\},\{y\})} = \mathbf{d}_{(x,y)} \quad \forall x, y \in I$$

Los términos diagonales de D(n,n) son nulos, puesto que, si d_{ij} es una distancia : $\mathbf{d}_{(\{x\},\{x\})} = \mathbf{d}_{(x,x)} = 0 \quad \forall x \in I$

Segunda etapa :

Buscamos en la tabla D(n,n) el término extra-diagonal mínimo, es decir el valor $\mathbf{d}_{(\{x\},\{y\})} = \mathbf{d}_{(x,y)}$ mínimo.

Formamos una nueva clase que reagrupa esos dos objetos: $\{x\}$ y $\{y\}$.

Iteración :

Se recomienza a partir de la primera etapa, pero ahora sólo con n - 1 objetos a comparar, puesto que una clase contiene ahora dos objetos.

Para calcular la Tabla D'(n-1, n-1) correspondiente a la nueva situación debemos darnos un criterio para calcular la distancia entre una clase que contiene dos objetos y las clases restantes que sólo contienen un objeto.

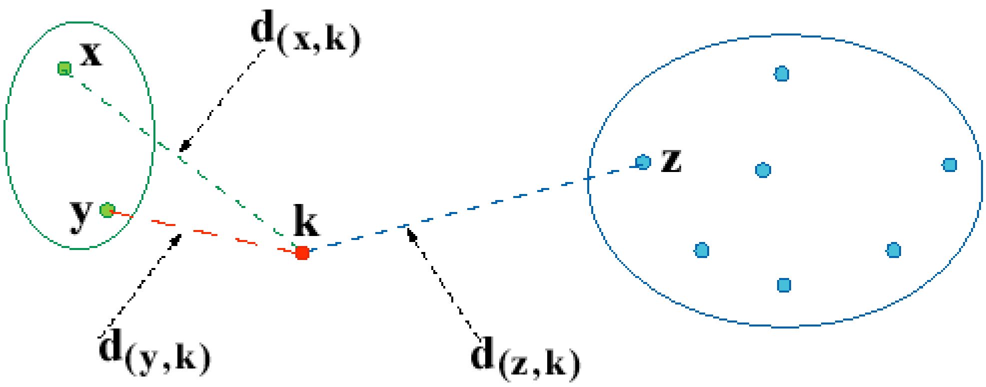
La estrategia de agregación responde a ese problema...

La estrategia del «vecino más cercano» consiste en elegir como distancia entre la clase $\{x;y\}$ y la clase $\{k\}$ la más pequeña de las dos distancias siguientes:

$$\mathbf{d}_{(\{x\},\{k\})} \quad \text{o bien} \quad \mathbf{d}_{(\{y\},\{k\})}$$

En cada etapa t de iteración del proceso de agregación por el método del «vecino más cercano», la Tabla D(n-t, n-t) es construida con la siguiente distancia ultramétrica :

$$d_{(\{x,y\},\{k\})} = Min\left(d_{\{x,k\}}; d_{\{y,k\}}\right)$$



✔ **Ventaja del método :** **simplicidad de cálculo. No requiere el cálculo de la matriz D_t(n-t, n-t) en cada etapa de agregación.**

✔ **Inconveniente del método :** **tiene tendencia a producir un «efecto de encadenamiento»**

INDICE

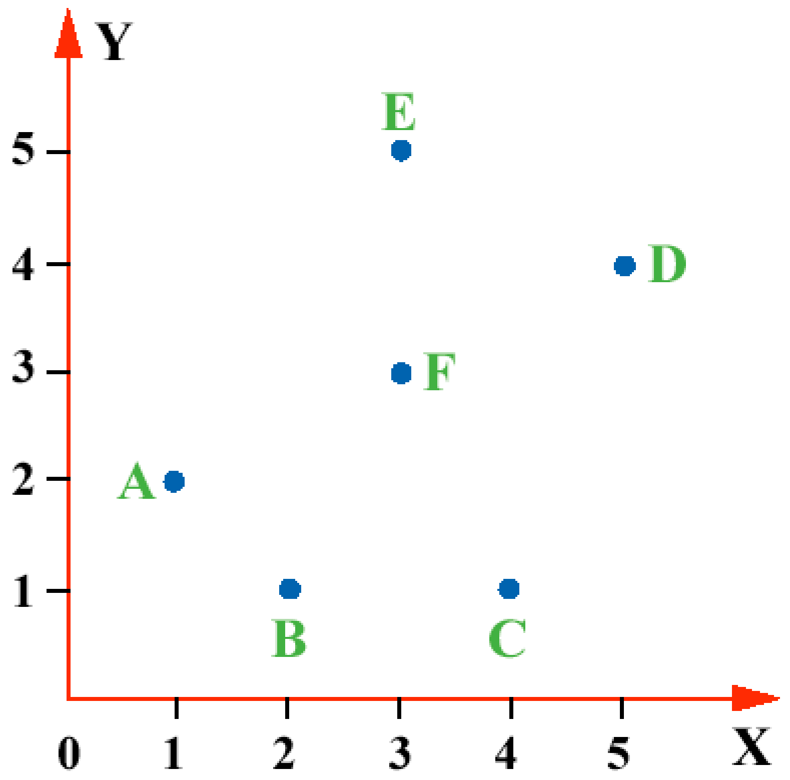
3.3. El método del «vecino más cercano» : un ejemplo numérico

a) Tabla de Datos y Representación gráfica en R^2

Tabla T(n,p)

	X	Y
A	1	2
B	2	1
C	4	1
D	5	4
E	3	5
F	3	3

Representación gráfica



INDICE

b) Primera agregación

Utilizando la distancia euclidiana,

$$d_{(i,j)} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad \forall i = 1,\ldots,n \ ; \ \forall j = 1,\ldots,n$$

podemos calcular la matriz $D_1(6, 6)$ siguiente :

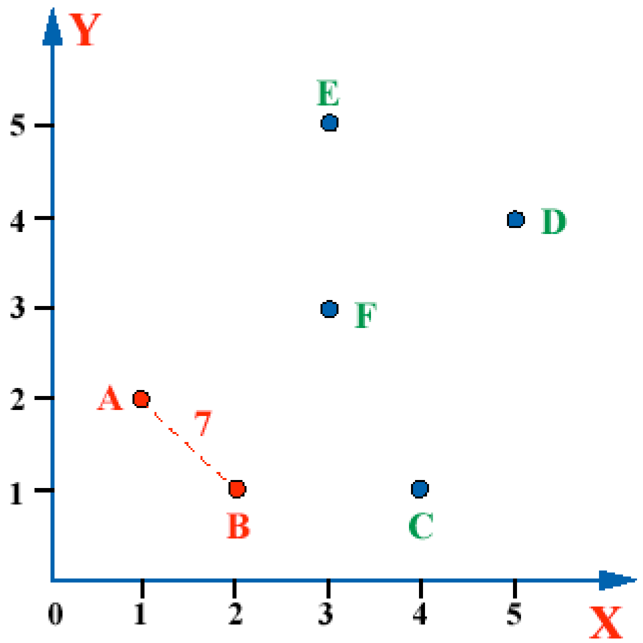
	A	B	C	D	E	F
A	0	1,41	3,16	4,47	3,61	2,24
B	1,41	0	2,00	4,24	4,12	2,24
C	3,16	2,00	0	3,16	4,12	2,24
D	4,47	4,24	3,16	0	2,24	2,24
E	3,61	4,12	4,12	2,24	0	2,00
F	2,24	2,24	2,24	2,24	2,00	0

Primera agregación : Como la distancia más pequeña se verifica entre los objetos A y B, formamos la clase {A, B}.

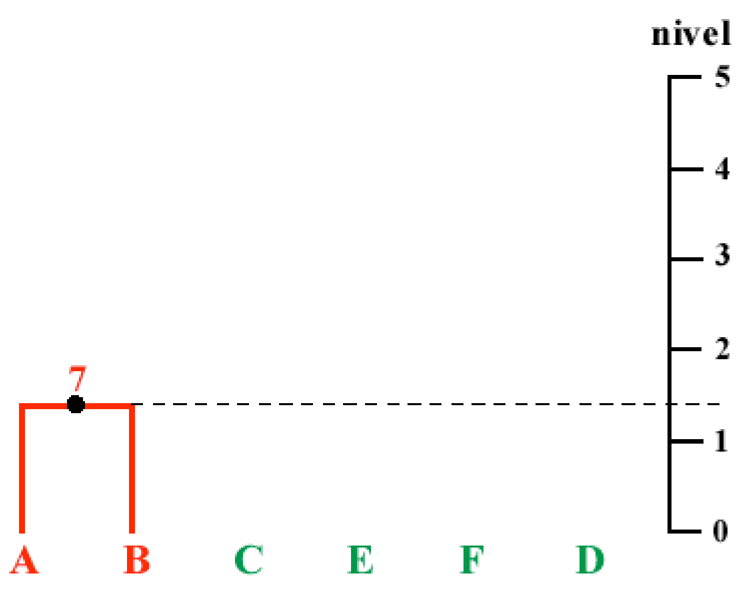
Nudo	Nivel	Primogénito	Benjamin	Peso
7	1,41	A	B	2

INDICE

Representación gráfica de la primera agregación



Dendrograma



c) Segunda agregación

Utilizando la distancia ultramétrica del «vecino más cercano», calculamos la tabla $D_2(5, 5)$ siguiente :

	(A,B)	C	D	E	F
(A,B)	0	2,00	4,24	3,61	2,24
C	2,00	0	3,16	4,12	2,24
D	4,24	3,16	0	2,24	2,24
E	3,61	4,12	2,24	0	2,00
F	2,24	2,24	2,24	2,00	0

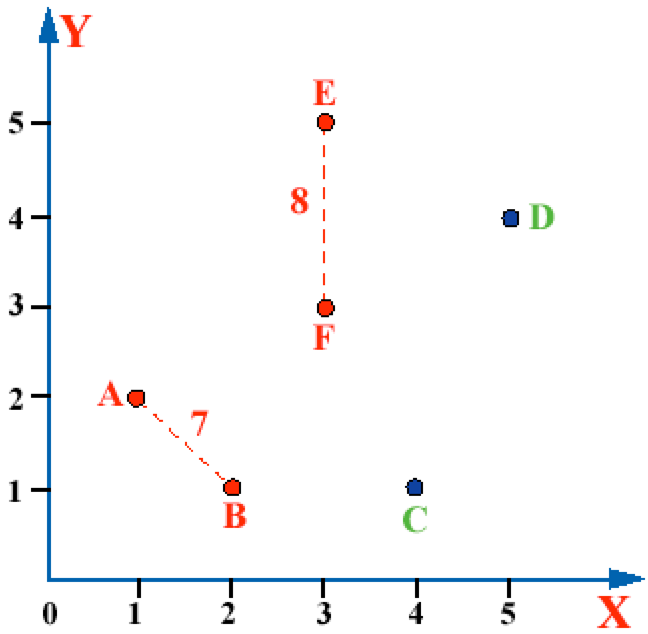
Segunda agregación : Dos pares de objetos presentan la distancia más pequeña :(E , F) y ((A, B), C). Formamos primero la clase{E, F}.

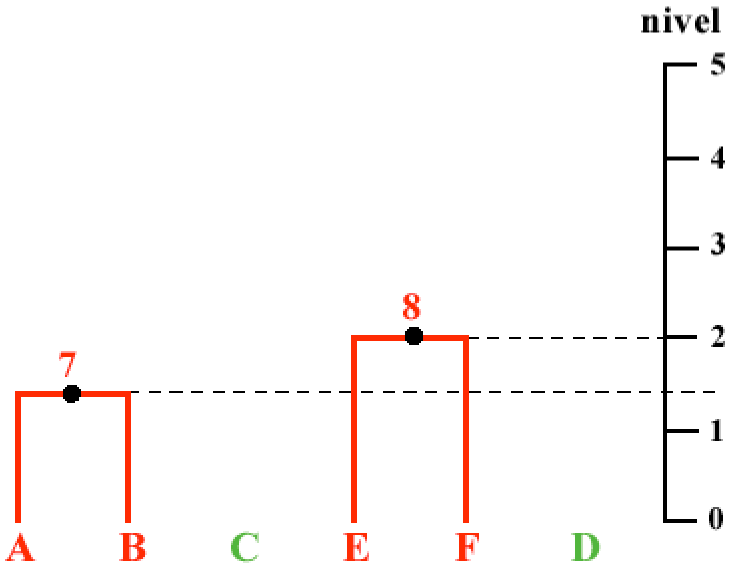
Nudo	Nivel	Primogénito	Benjamin	Peso
7	1,41	A	B	2
8	2,00	E	F	2

INDICE

Representación gráfica de la segunda agregación

Dendrograma





d) Tercera agregación

Utilizando la distancia ultramétrica del «vecino más cercano», calculamos la tabla $D_3(4, 4)$ siguiente :

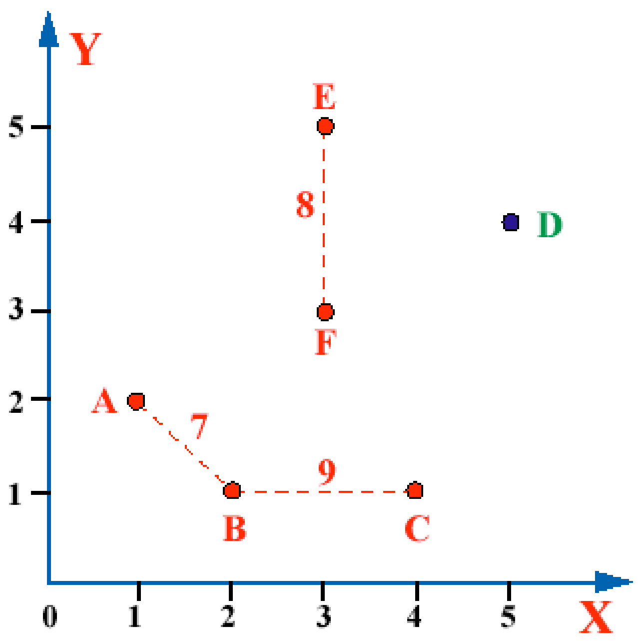
	(A,B)	C	D	(E,F)
(A,B)	0	2,00	4,24	2,24
C	2,00	0	3,16	2,24
D	4,24	3,16	0	2,24
(E,F)	2,24	2,24	2,24	0

Tercera agregación : Como la distancia más pequeña se verifica entre los objetos (A, B) y (C), formamos la clase **{A, B, C}**.

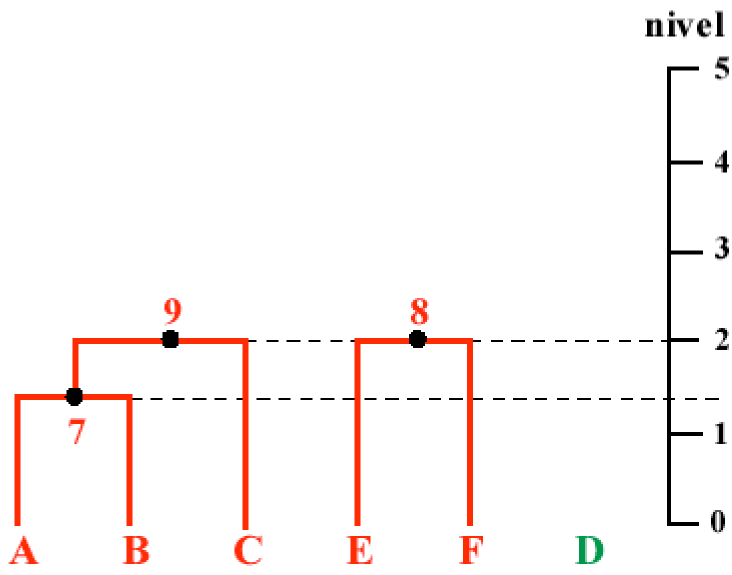
Nudo	Nivel	Primogénito	Benjamin	Peso
7	1,41	A	B	2
8	2,00	E	F	2
9	2,00	AB	C	3

INDICE

Representación gráfica de la tercera agregación



Dendrograma



e) Cuarta agregación

Utilizando la distancia ultramétrica del «vecino más cercano», calculamos la tabla $D_4(3, 3)$ siguiente :

	(A,B,C)	D	(E,F)
(A,B,C)	0	3,16	2,24
D	3,16	0	2,24
(E,F)	2,24	2,24	0

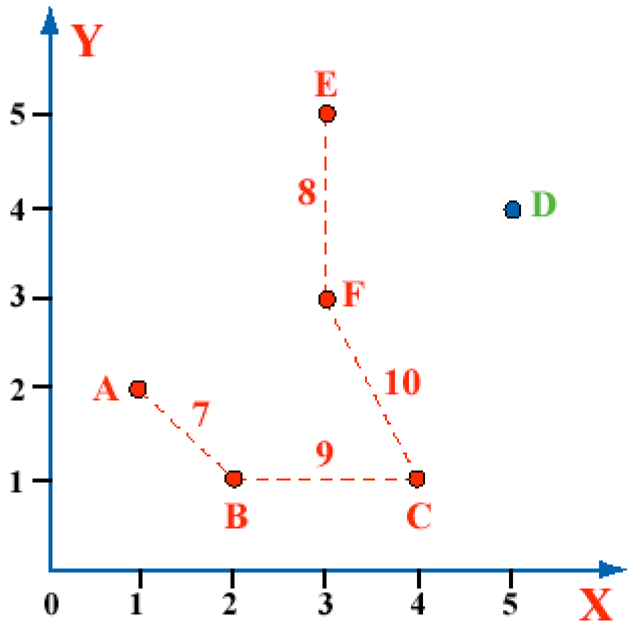
Cuarta agregación : Dos pares de objetos presentan la distancia más pequeña :({A, B, C} y {E, F}) y ({E, F} y {D}). Formamos primero la clase{A, B, C, E, F} :

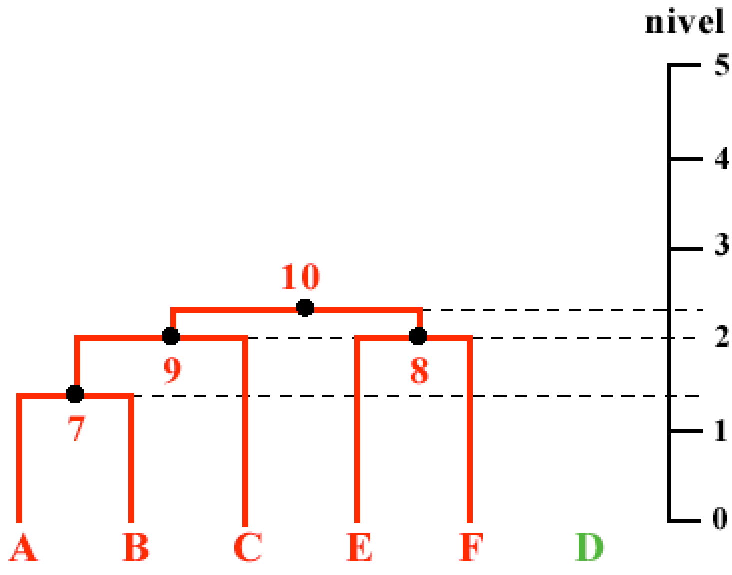
Nudo	Nivel	Primogénito	Benjamin	Peso
7	1,41	A	B	2
8	2,00	E	F	2
9	2,00	AB	C	3
10	2,24	ABC	EF	5

INDICE

Representación gráfica de la cuarta agregación

Dendrograma





f) Quinta agregación

Utilizando la distancia ultramétrica del «vecino más cercano», calculamos la tabla $D_5(2, 2)$ siguiente :

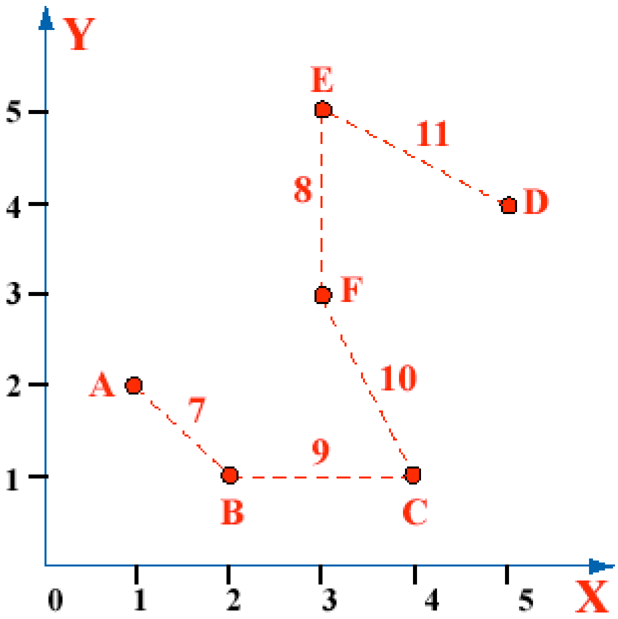
	(A,B,C,E,F)	D
(A,B,C,E,F)	0	2,24
D	2,24	0

Quinta agregación : **Se agrupan por último los objetos {A,B,C, E, F} y {D}, formando la clase {A,B,C,E,F,D} que reúne todos los objetos de T(n, p)**

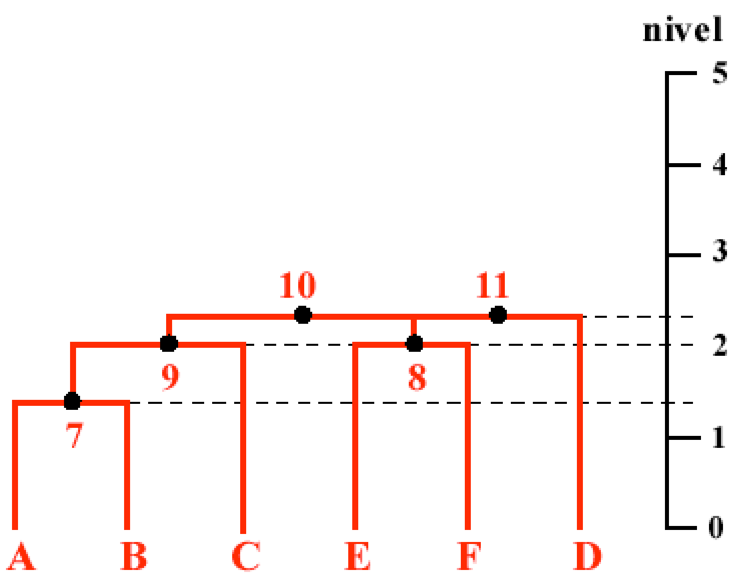
Nudo	Nivel	Primogénito	Benjamin	Peso
7	1,41	A	B	2
8	2,00	E	F	2
9	2,00	AB	C	3
10	2,24	ABC	EF	5
11	2,24	ABCEF	D	6

INDICE

Representación gráfica de la quinta agregación



Dendrograma final



g) Resultados de la clasificación

Descripción de las clases encajadas sucesivas

Nudo	Nivel	Primogénito	Benjamin	Peso
7	1,41	A	B	2
8	2,00	E	F	2
9	2,00	AB	C	3
10	2,24	ABC	EF	5
11	2,24	ABCEF	D	6

Histograma de índices de nivel

Num.	Primog.	Benjamin	Efec.	Peso	Indice	Histogr. de los índices de nivel
7	A	B	2	2.00	1.41	***************
8	E	F	2	2.00	2.00	********************
9	7	C	3	3.00	2.00	********************
10	9	8	5	5.00	2.24	*************************
11	10	D	6	6.00	2.24	*************************