



**INSTITUTO POLITÉCNICO
NACIONAL**



**ESCUELA SUPERIOR DE
CÓMPUTO**

DATA MINING

Proyecto No. 4. Clúster

Grupo: 3CV15

profesora: Ocampo Botello Fabiola

Equipos: 4 y 10

Integrantes equipo 4:

- Flores Ponce Alan Marcelo
- García Cruz Octavio Arturo
- Sampayo Hernández Mauro

Integrantes equipo 10:

- Alanis Garduño Mireya
- Rios Rebollar Victor Hugo

PARTE 1. Diseñar una matriz con 6 puntuaciones (x, y) y desarrollar el proceso paso a paso de la creación de grupos aplicando las tres siguientes técnicas (*Inserte una portada indicando el tema que aborda*)

1. Enlace simple
2. Enlace promedio
3. Enlace completo



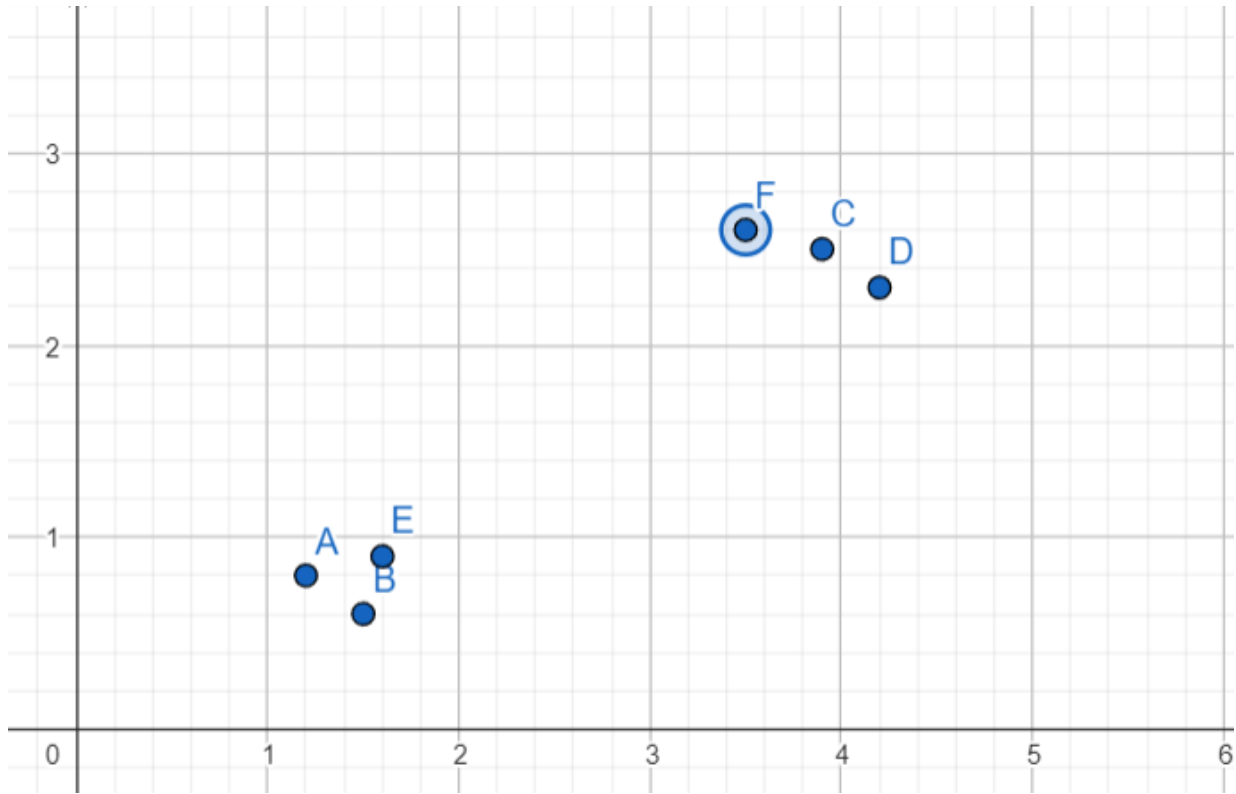
ENLACE SIMPLE

PASO 1:

Matriz de puntuaciones (x,y)

	x	y
A	1.2	0.8
B	1.5	0.6
C	3.9	2.5
D	4.2	2.3
E	1.6	0.9
F	3.5	2.6

Comenzamos graficando los vectores:

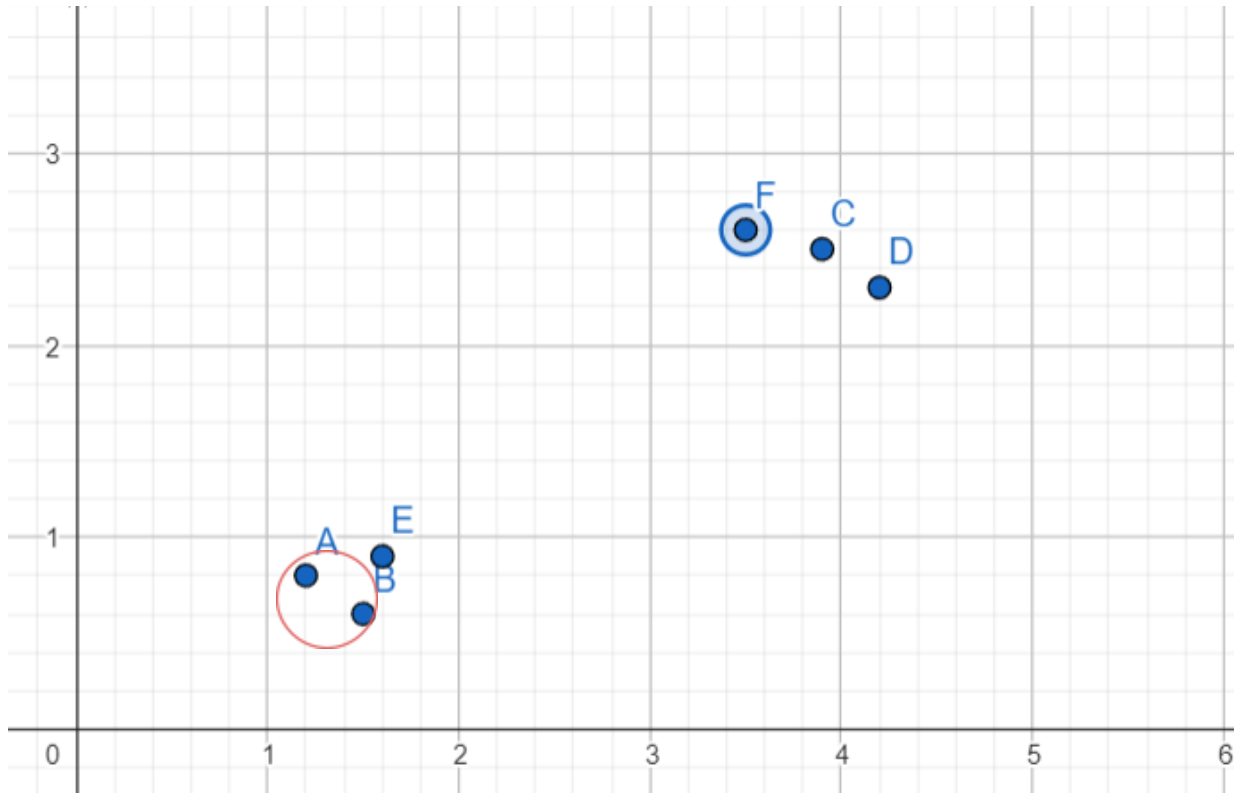


usando la distancia euclidiana pasamos a construir la primera matriz de distancias:

$$d(p_1, p_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

	A	B	C	D	E	F
A	0					
B	0.2828	0				
C	3.7890	3.2455	0			
D	3.9925	3.4460	0.3162	0		
E	0.5385	0.9487	2.8302	2.39	0	
F	3.7270	3.2211	0.7071	2.71	1.51	0

Entonces, AB será nuestro primer cluster



PASO 2

Una vez encontrado el valor más pequeño, rehacemos la matriz de distancias considerando dicho valor.

En este caso utilizaremos la fórmula $MIN[dist(A, B), C]$
 $MIN[(A, C), (B, C)] = MIN(3.7890, 3.2455)$
 $= 3.2455$

En este caso utilizaremos la fórmula $MIN[dist(A, B), D]$
 $MIN[(A, D), (B, D)] = MIN(3.9925, 3.4460)$
 $= 3.4460$

En este caso utilizaremos la fórmula $MIN[dist(A, B), E]$
 $MIN[(A, E), (B, E)] = MIN(0.5385, 0.9487)$
 $= 0.5385$

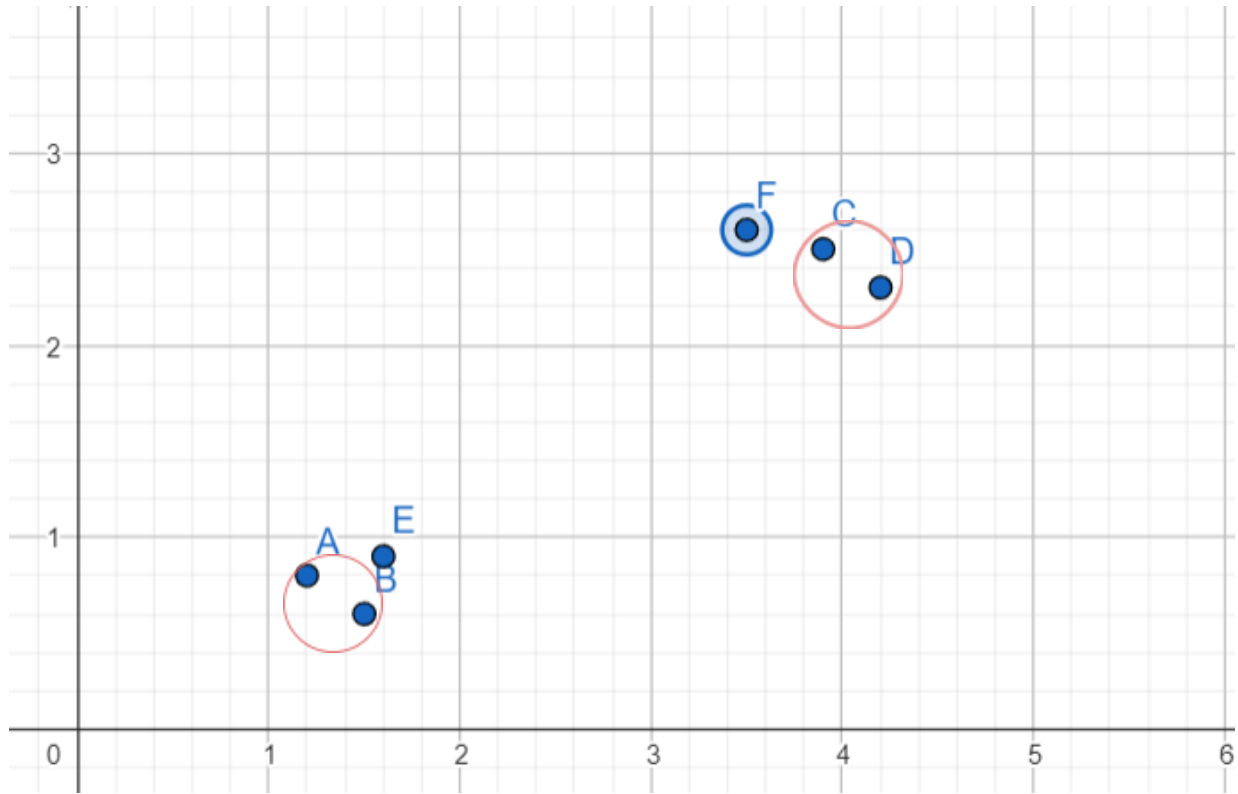
En este caso utilizaremos la fórmula $MIN[dist(A, B), F]$
 $MIN[(A, F), (B, F)] = MIN(3.7270, 3.2211)$
 $= 3.2211$

En este caso utilizaremos la fórmula $MIN[dist(C, D), AB]$
 $MIN[(C, AB), (D, AB)] = MIN(3.7890, 3.2455)$
 $= 3.2455$

En este caso utilizaremos la fórmula $MIN[dist(C, D), E]$
 $MIN[(C, E), (D, E)] = MIN(2.8302, 2.39)$
 $= 2.39$

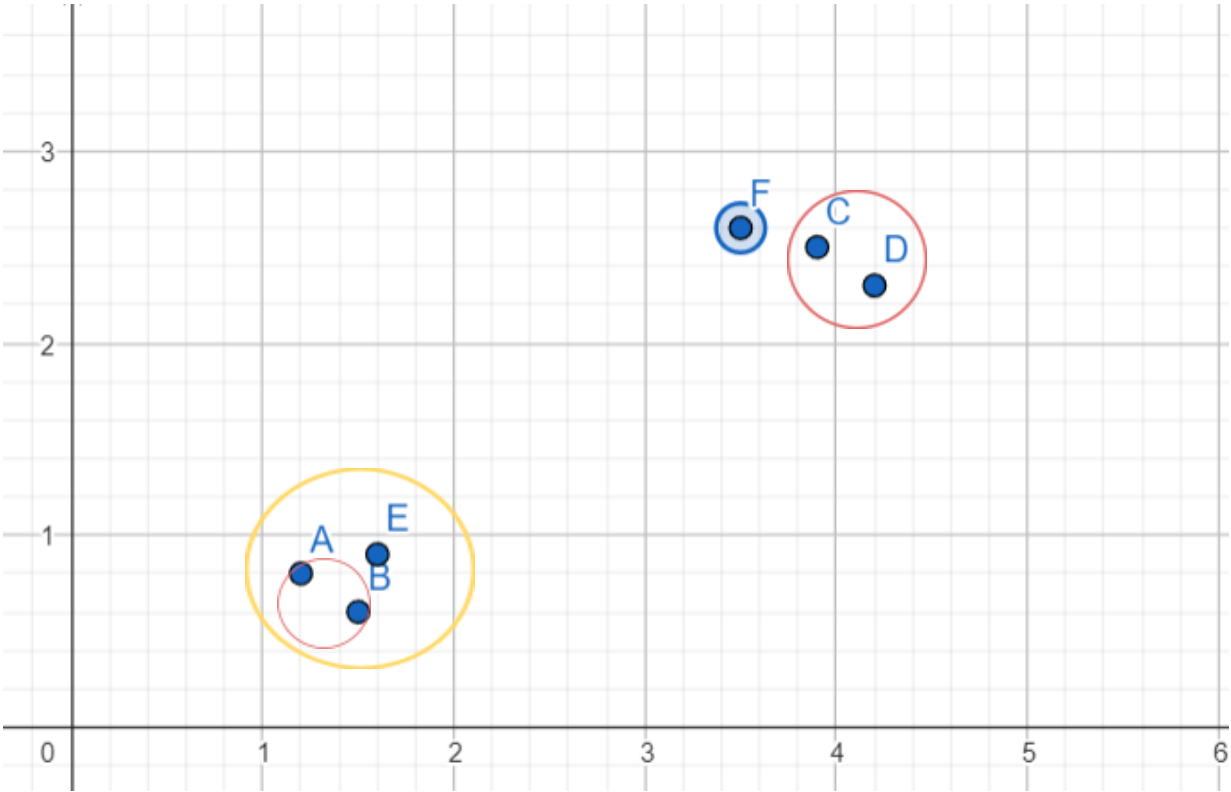
En este caso utilizaremos la fórmula $MIN[dist(C, D), F]$
 $MIN[(C, F), (D, F)] = MIN(0.7172, 2.71)$
 $= 0.7172$

	AB	C	D	E	F
AB	0				
C	3.2455	0			
D	3.4460	0.3162	0		
E	0.5385	2.39	2.39	0	
F	3.2211	0.7071	2.71	1.51	0



De nuevo, rehacemos la matriz:

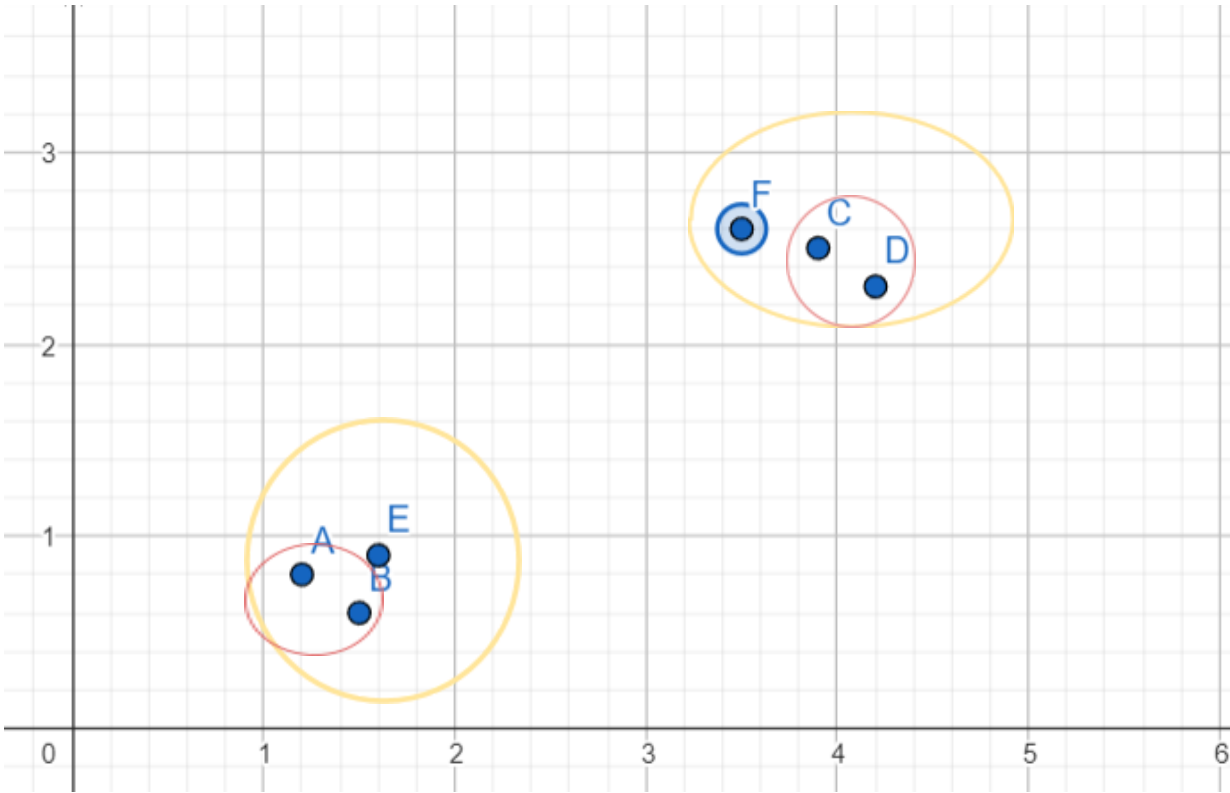
	AB	CD	E	F
AB	0			
CD	3.2455	0		
E	0.5385	2.39	0	
F	3.2211	0.7071	1.51	0



Distancia de CD con (AB,E)= $MIN[dist(AB,E),CD]$
 $MIN[dist(AB,CD),(E,CD)] = MIN[(3.7192,2.6101)]$
 = 2.6101

Distancia de F con (AB,E)= $MIN[dist(AB,E),F]$
 $MIN[dist(AB,F),(E,F)] = MIN[(3.4740,1.51)]$
 = 1.51

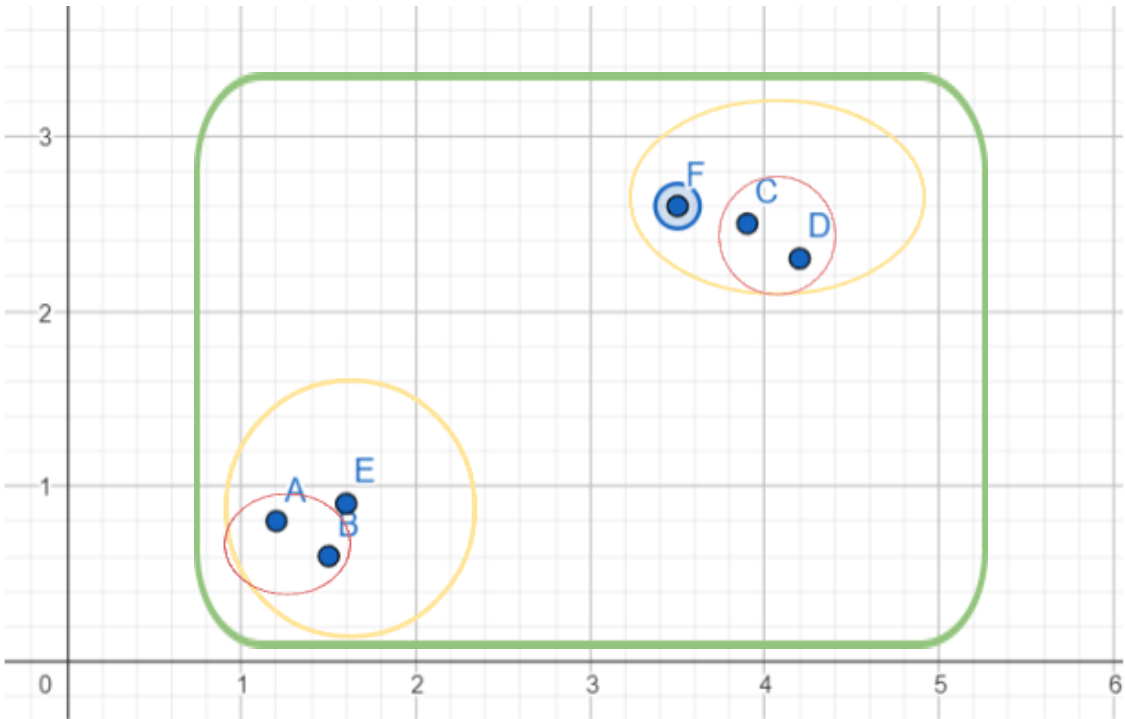
	ABE	CD	F
ABE	0		
CD	2.6101	0	
F	1.51	0.7071	0



Calculamos nuevamente:
Distancia de ABE con (CD,F)= $MIN[dist(CD,F), ABE]$
 $MIN[dist(CD, ABE), (F, ABE)] = MIN[3.1646, 2.495]$
= 2.495

	ABE	CDF
ABE	0	
CDF	2.495	0

Finalmente, formamos el último cluster:





ENLACE PROMEDIO

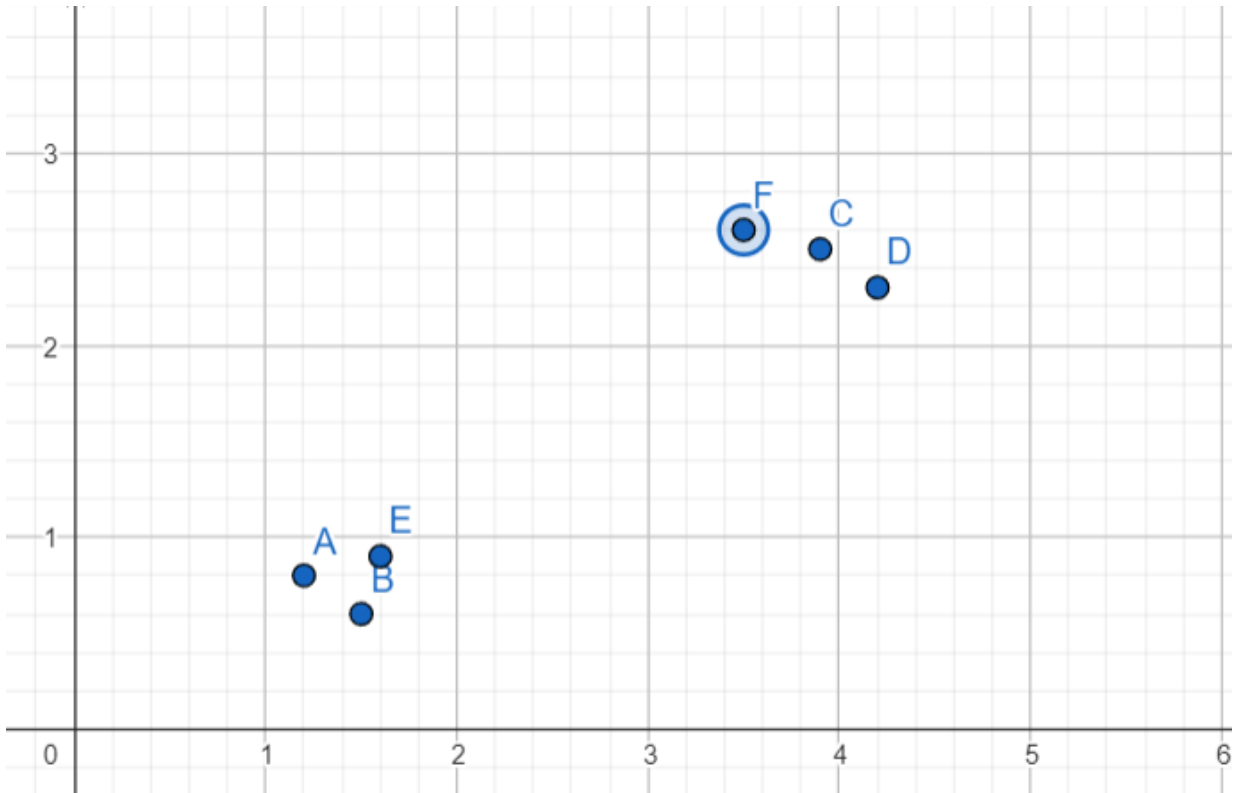
PASO 1:

Matriz de puntuaciones (x,y)

	x	y
A	1.2	0.8
B	1.5	0.6
C	3.9	2.5
D	4.2	2.3
E	1.6	0.9
F	3.5	2.6

Seguiremos utilizando la matriz generada en el enlace simple.

Comenzamos graficando los vectores:



usando la distancia euclidiana pasamos a construir la primera matriz de distancias:

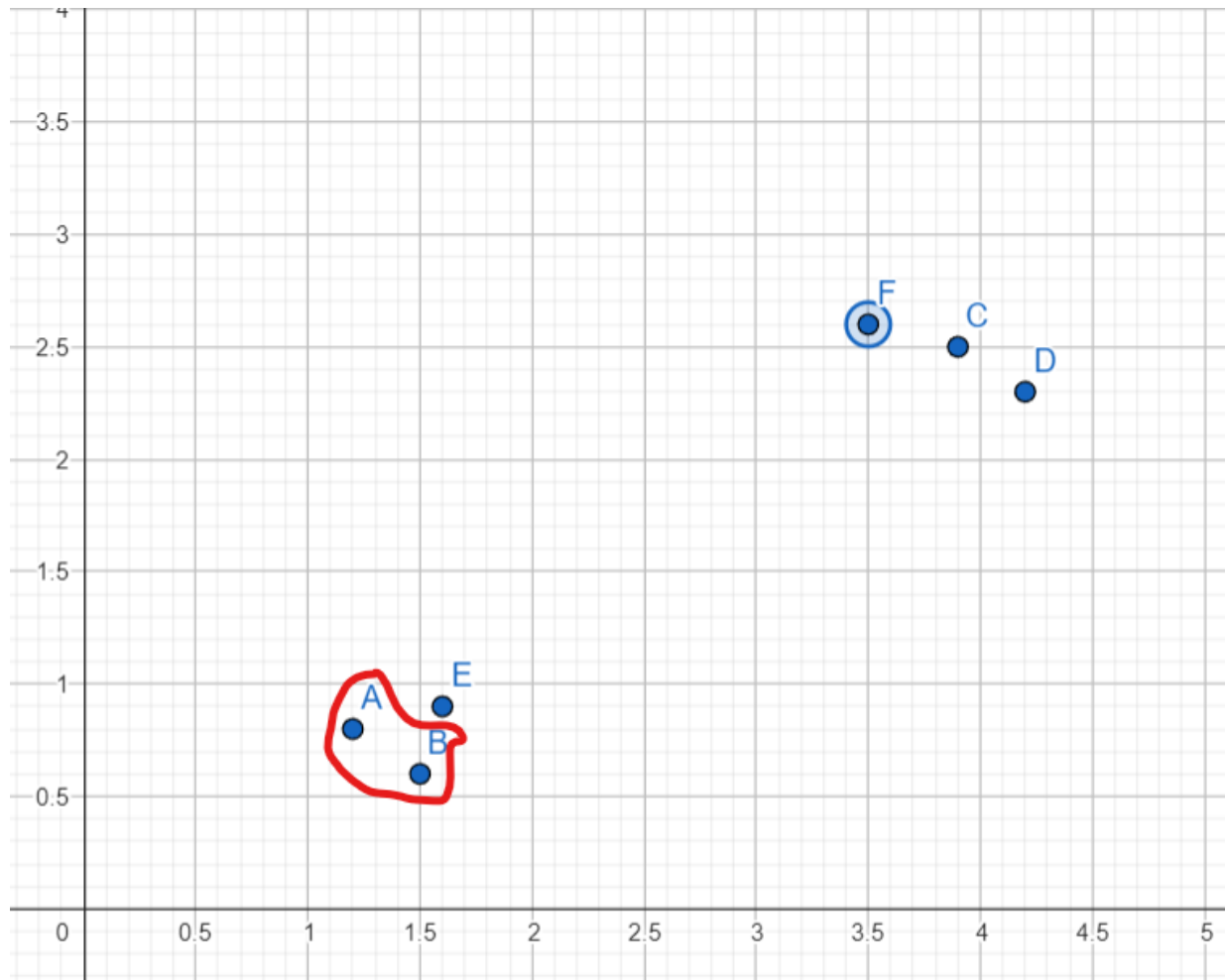
$$d(a,b) = \sqrt{((x_2 - x_1)^2 + (y_2 - y_1)^2)}$$

	A	B	C	D	E	F
A	0					
B	0.2828	0				
C	3.7890	3.2455	0			
D	3.9925	3.4460	0.3162	0		
E	0.5385	0.9487	2.8302	2.39	0	
F	3.7270	3.2211	0.7071	2.71	1.51	0

Buscamos la distancia más corta entre los puntos, que como podemos ver se encuentra en:

$$d(a,b)=0.2828$$

y este punto (a,b) será nuestro primer cluster generado



A continuación, realizaremos los cálculos para obtener la distancia entre AB y los demás puntos

$$\begin{aligned} \text{Distancia de C con AB} &= \text{AVG}[\text{dist}(AB), C] \\ \text{AVG}[\text{dist}(A, C), (B, C)] &= \text{AVG}[(3.7890, 3.2455)] \\ &= 1/2(3.7890 + 3.2455) \\ &= 3.5172 \end{aligned}$$

$$\begin{aligned} \text{Distancia de D con AB} &= \text{AVG}[\text{dist}(AB), D] \\ \text{AVG}[\text{dist}(A, D), (B, D)] &= \text{AVG}[(3.9925, 3.4460)] \\ &= 1/2(3.9925 + 3.4460) \\ &= 3.7192 \end{aligned}$$

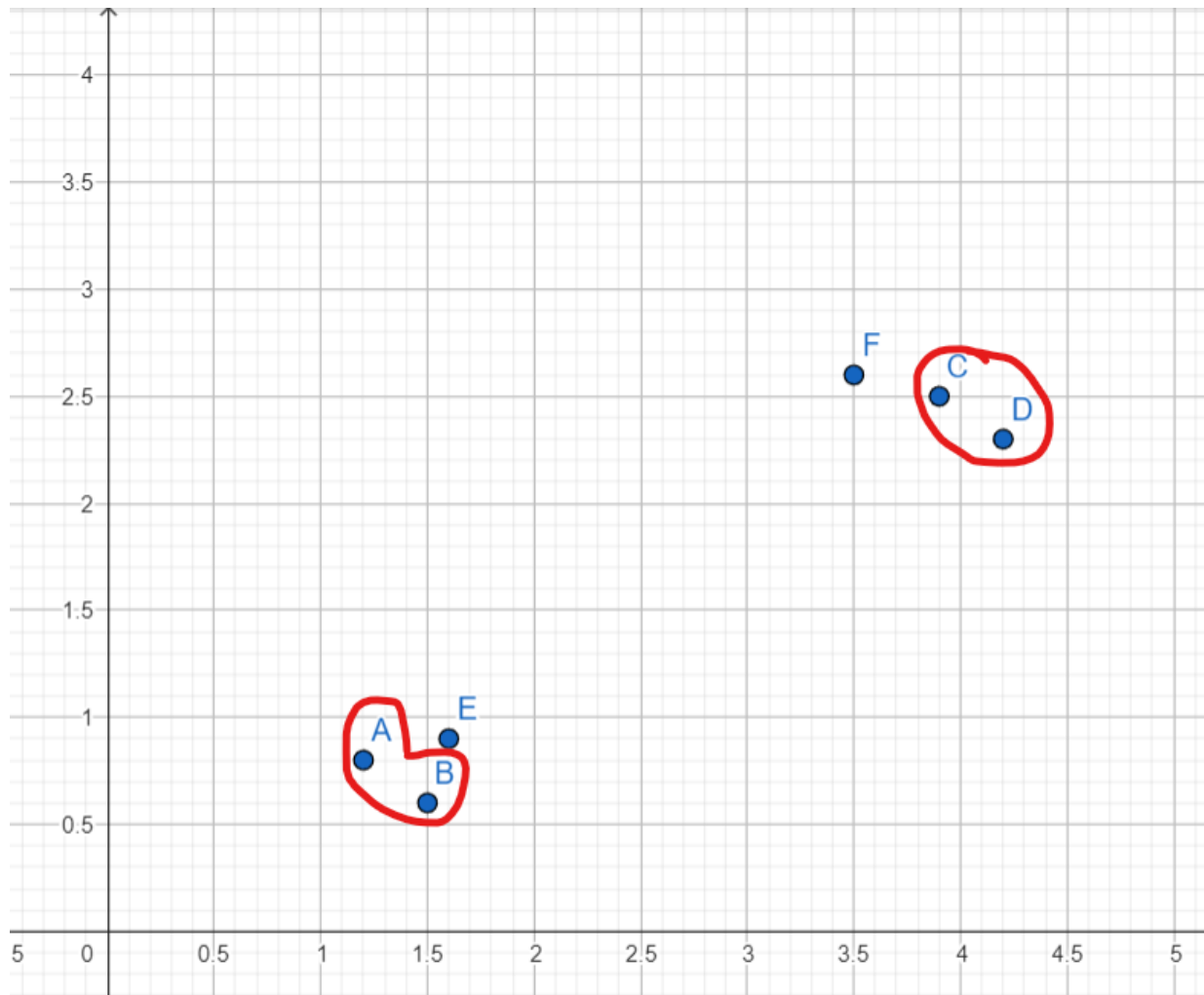
Distancia de E con AB= $AVG[dist(AB), E]$
 $AVG[dist(A, E), (B, E)] = AVG[(0.5385, 0.9487)]$
 $= 1/2(0.5385 + 0.9487)$
 $= 0.7436$

Distancia de F con AB= $AVG[dist(AB), F]$
 $AVG[dist(A, F), (B, F)] = AVG[(3.7270, 3.2211)]$
 $= 1/2(3.7270 + 3.2211)$
 $= 3.4740$

Recalculamos la matriz con los nuevos datos que acabamos de calcular y agrupando con el cluster creado

	AB	C	D	E	F
AB	0				
C	3.5172	0			
D	3.7192	0.3162	0		
E	0.7436	2.8302	2.39	0	
F	3.4740	0.7071	2.71	1.51	0

En este caso la distancia más corta está entre los puntos CD, por lo tanto estos puntos crean otro cluster, por lo tanto, ya tenemos 2 clusters agrupados en nuestro conjunto



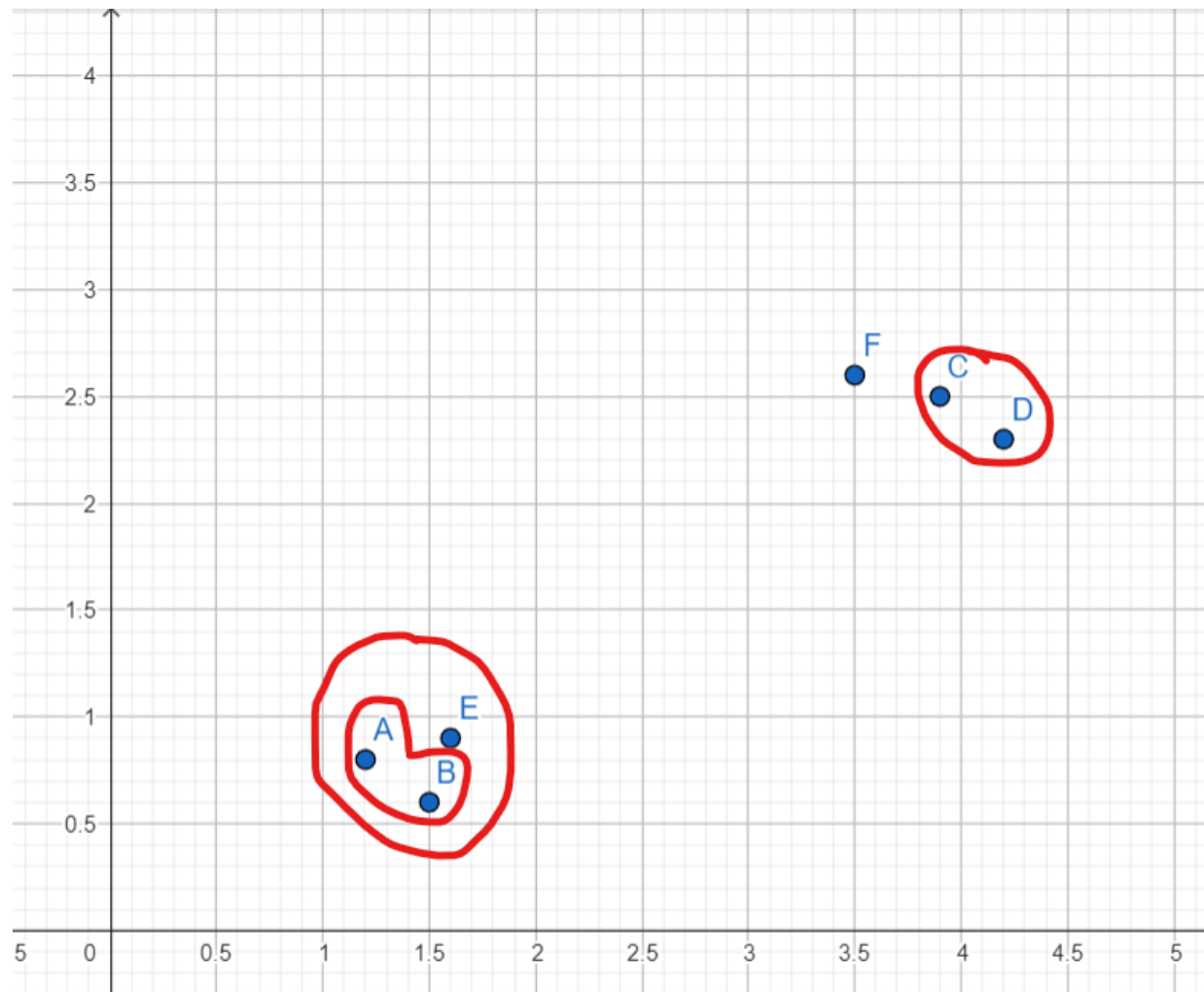
volvemos aplicar el proceso de calcular el promedio de este nuevo cluster con los demás vectores en la matriz

Distancia de E con CD= $AVG[dist(CD), E]$
 $AVG[dist(C, E), (D, E)] = AVG[(2.8302, 2.39)]$
 $= 1/2(2.8302 + 2.39)$
 $= 2.6101$

Distancia de F con CD= $AVG[dist(CD), F]$
 $AVG[dist(C, F), (D, F)] = AVG[(0.7071, 2.71)]$
 $= 1/2(0.7071 + 2.71)$
 $= 1.7085$

	AB	CD	E	F
AB	0			
CD	3.7192	0		
E	0.7436	2.6101	0	
F	3.4740	1.7085	1.51	0

De nuevo buscamos la distancia más pequeña, y podemos ver que se encuentra en la intersección de AB con E y agrupamos el cluster nuevo



ahora obtenemos los promedios para calcular la nueva matriz

Distancia de CD con (AB,E)= $AVG[dist(AB,E),CD]$
 $AVG[dist(AB,CD), (E,CD)] = AVG[(3.7192, 2.6101)]$
 $= 1/2(3.7192 + 2.6101)$
 $= 3.1646$

Distancia de F con (AB,E)= $AVG[dist(AB,E),F]$
 $AVG[dist(AB,F), (E,F)] = AVG[(3.4740, 1.51)]$
 $= 1/2(3.4740 + 1.51)$
 $= 2.495$

	ABE	CD	F
ABE	0		
CD	3.1646	0	
F	2.495	1.7085	0

Seleccionamos de nueva cuenta la distancia menor en la tabla,
 podemos ver que ocurre en la intersección entre por CDF lo tanto
 este sería un nuevo cluster EF

Distancia de ABE con (CD,F)= $AVG[dist(CD,F),ABE]$
 $AVG[dist(CD,ABE), (F,ABE)] = AVG[(3.1646, 2.495)]$
 $= 1/2(3.1646 + 2.495)$
 $= 2.8298$

	ABE	CDF
ABE	0	
CDF	2.8298	0

Con este último valor formamos nuestro último cluster agrupando
 todos los vectores





ENLACE COMPLETO

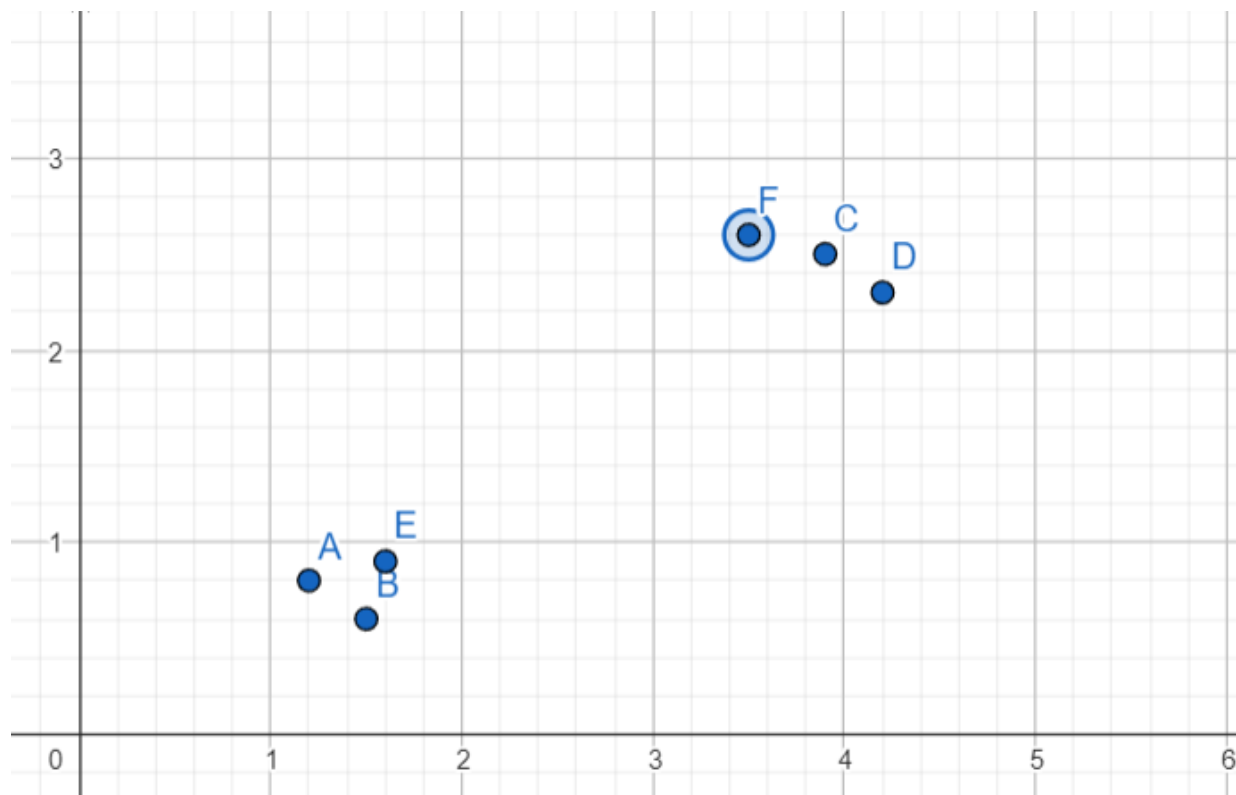
PASO 1:

Matriz de puntuaciones (x,y)

	x	y
A	1.2	0.8
B	1.5	0.6
C	3.9	2.5
D	4.2	2.3
E	1.6	0.9
F	3.5	2.6

Seguiremos utilizando la matriz generada en los 2 anteriores ejemplos

Comenzamos graficando los vectores:



usando la distancia euclidiana pasamos a construir la primera matriz de distancias:

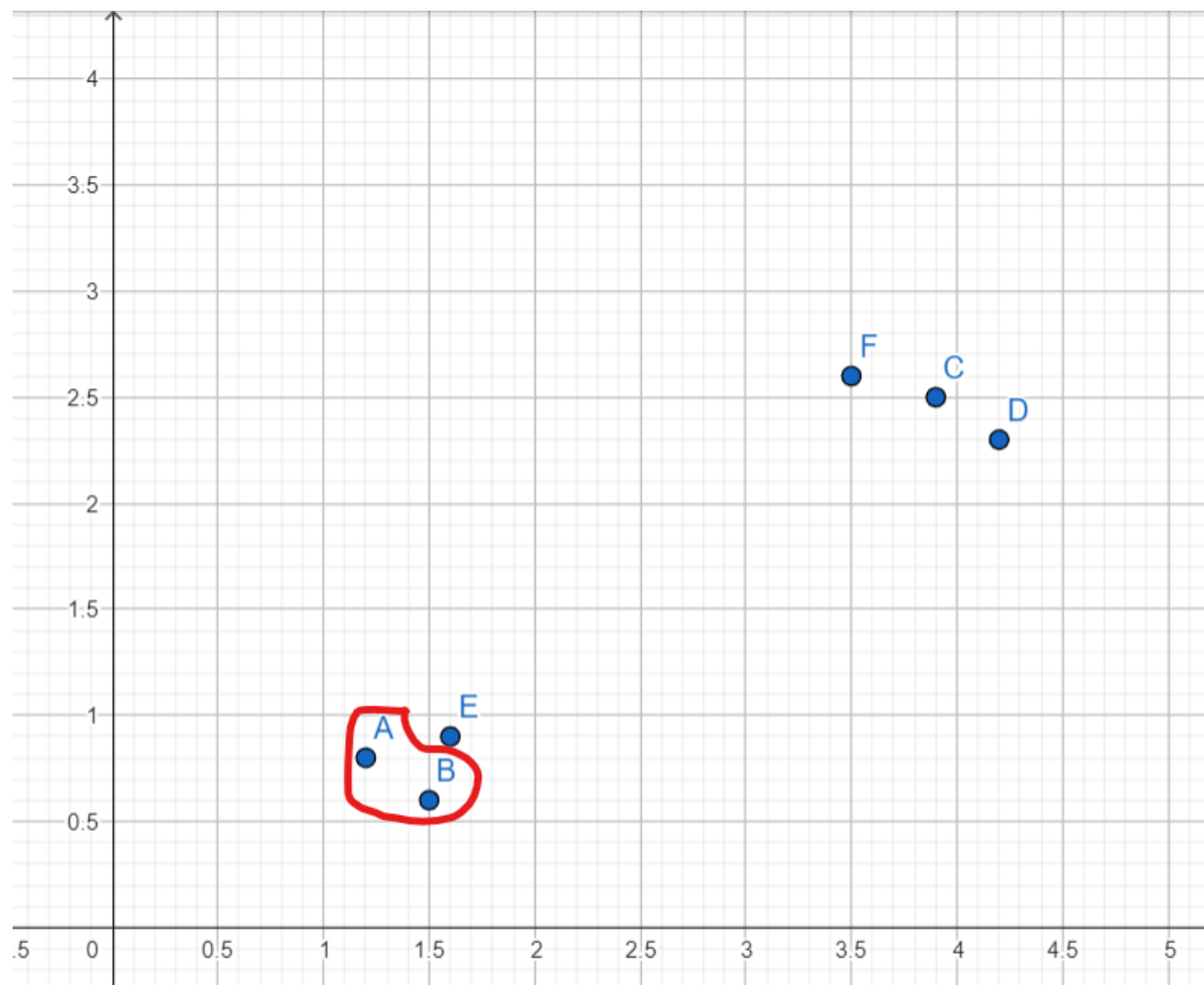
$$d(a,b) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

	A	B	C	D	E	F
A	0					
B	0.2828	0				
C	3.7890	3.2455	0			
D	3.9925	3.4460	0.3162	0		
E	0.5385	0.9487	2.8302	2.39	0	
F	3.7270	3.2211	0.7071	2.71	1.51	0

Buscamos la distancia más corta entre los puntos, que como podemos ver se encuentra en:

$$d(a,b)=0.2828$$

y este punto (a,b) será nuestro primer cluster generado



En este caso utilizaremos la fórmula $MAX[dist(A, B), C]$
 $MAX[(A, C), (B, C)] = MAX(3.7890, 3.2455)$
 $= 3.7890$

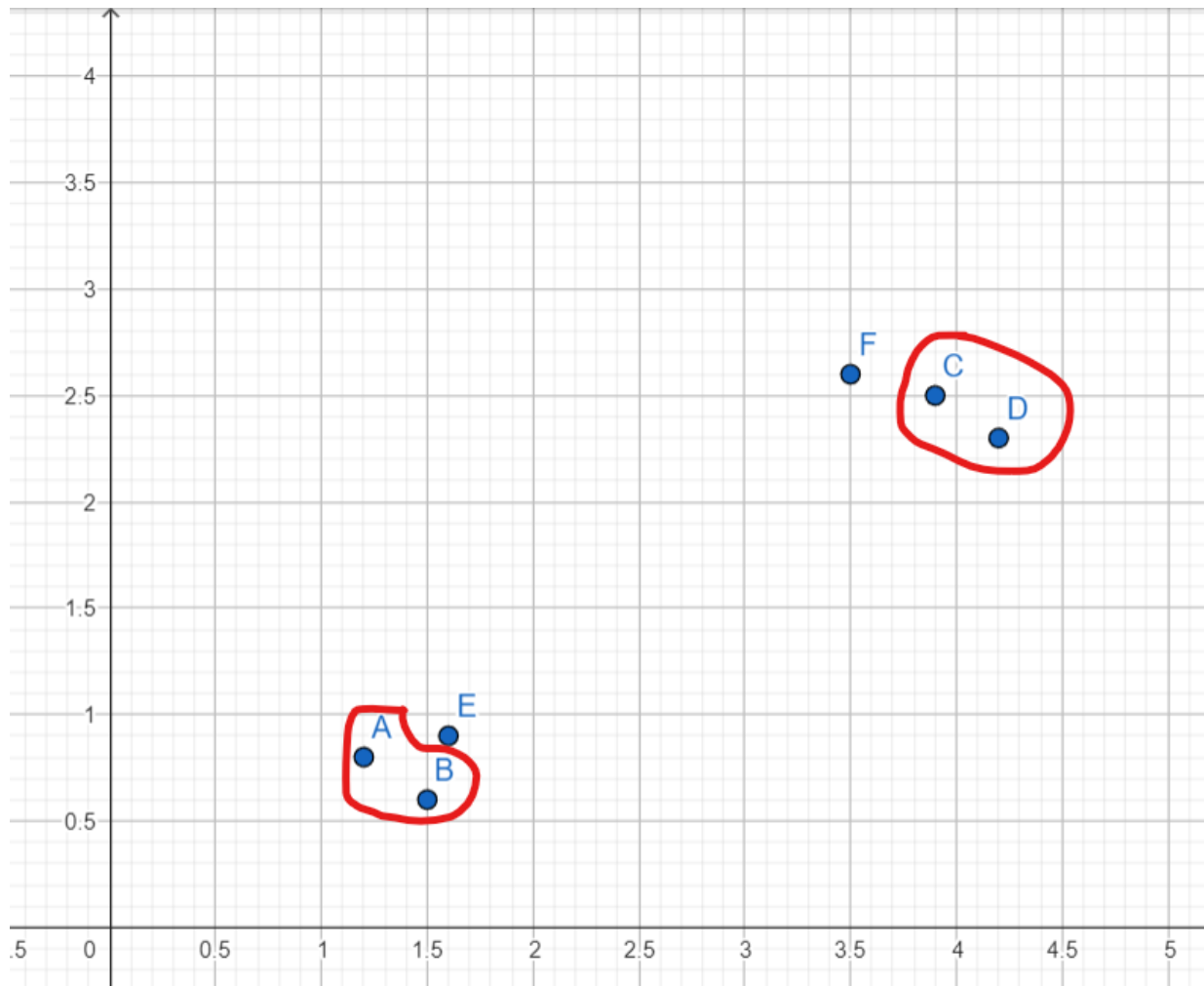
En este caso utilizaremos la fórmula $MAX[dist(A, B), D]$
 $MAX[(A, D), (B, D)] = MAX(3.9925, 3.4460)$
 $= 3.9925$

En este caso utilizaremos la fórmula $MAX[dist(A, B), E]$
 $MAX[(A, E), (B, E)] = MAX(0.5385, 0.9487)$
 $= 0.9487$

En este caso utilizaremos la fórmula $MAX[dist(A, B), F]$
 $MAX[(A, F), (B, F)] = MAX(3.7270, 3.2211)$
 $= 3.7270$

	AB	C	D	E	F
AB	0				
C	3.7890	0			
D	3.9925	0.3162	0		
E	0.9487	2.8302	2.39	0	
F	3.7270	0.7071	2.71	1.51	0

Ahora tenemos la distancia entre C Y D como la menor, por lo tanto CD es nuestro otro cluster



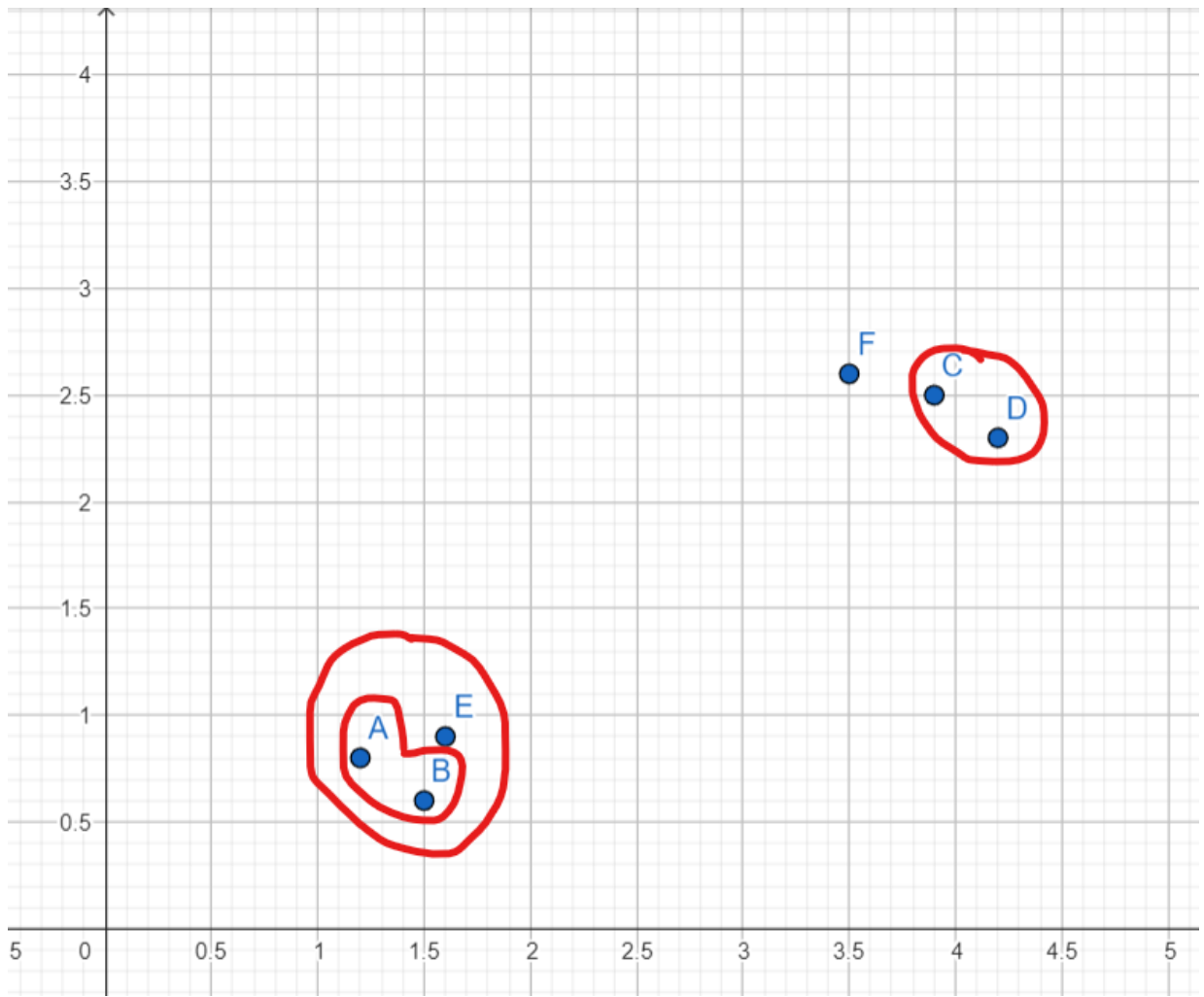
En este caso utilizaremos la fórmula $MAX[dist(C,D), AB]$
 $MAX[(C, AB), (D, AB)] = MAX(3.7890, 3.2455)$
 $= 3.7890$

En este caso utilizaremos la fórmula $MAX[dist(C,D), E]$
 $MAX[(C, E), (D, E)] = MAX(2.8302, 2.39)$
 $= 2.8302$

En este caso utilizaremos la fórmula $MAX[dist(C,D), F]$
 $MAX[(C, F), (D, F)] = MAX(0.7172, 2.71)$
 $= 2.71$

	AB	CD	E	F
AB	0			
CD	3.7890	0		
E	0.9487	2.8302	0	
F	3.7270	2.71	1.51	0

Seleccionamos el valor mínimo, el cual intersecciona en los puntos ABE

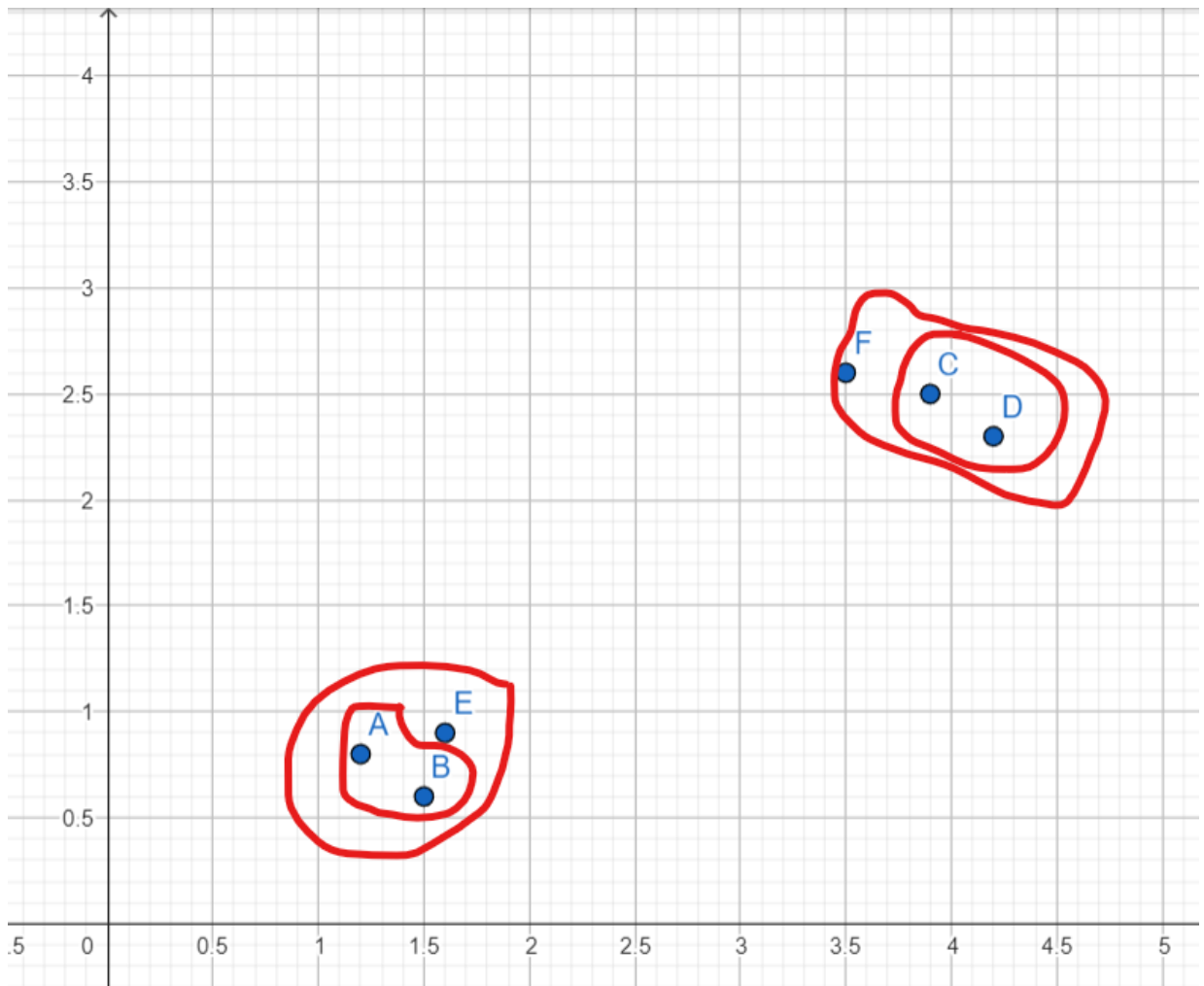


Distancia de CD con (AB,E)= $MAX[dist(AB,E), CD]$
 $MAX[dist(AB,CD), (E,CD)] = MAX[(3.7192, 2.6101)]$
 $= 3.7192$

Distancia de F con (AB,E)= $MAX[dist(AB,E), F]$
 $MAX[dist(AB,F), (E,F)] = MAX[(3.4740, 1.51)]$
 $= 3.4740$

	ABE	CD	F
ABE	0		
CD	3.7192	0	
F	3.4740	1.7085	0

De nuevo escogemos la distancia mínima, la cual intersecciona en los puntos CD y F



Distancia de ABE con (CD, F) = $MAX[dist(CD, F), ABE]$
 $MAX[dist(CD, ABE), (F, ABE)] = MAX[(3.1646, 2.495)]$
= 3.1646

	ABE	CDF
ABE	0	
CDF	2.8298	0

Con este último valor formamos nuestro último cluster agrupando todos los vectores



Con esto, ya tenemos agrupados todos los vectores en un cluster