

# **Μάθημα 8**

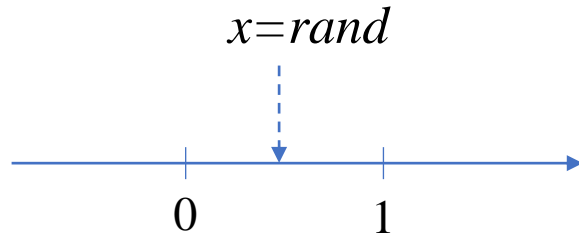
## **ΕΞΟΥΥΞΗ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΨΗΦΙΑΚΟ ΚΑΙ ΔΙΑΔΙΚΤΥΑΚΟ ΠΕΡΙΕΧΟΜΕΝΟ**

**Συσταδοποίηση Δεδομένων (Συνέχεια)**

# Αλγόριθμος c-Means

Τυχαία Επιλογή Αριθμών: Η Εντολή *rand*

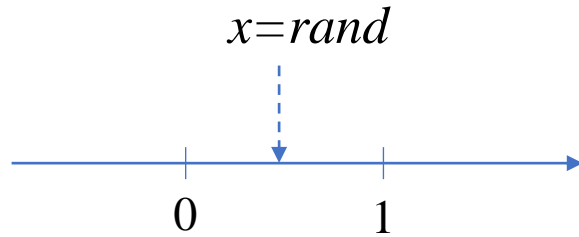
Επιλογή τυχαίας τιμής στο διάστημα [0 1]



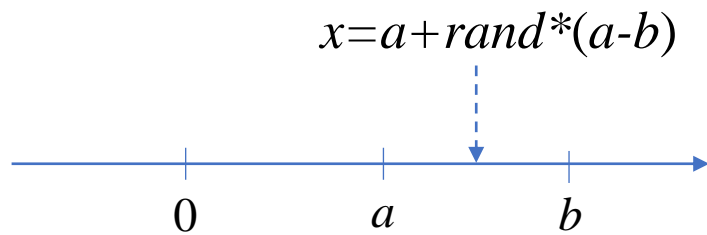
# Αλγόριθμος c-Means

## Τυχαία Επιλογή Αριθμών: Η Εντολή *rand*

Επιλογή τυχαίας τιμής στο διάστημα  $[0 \ 1]$



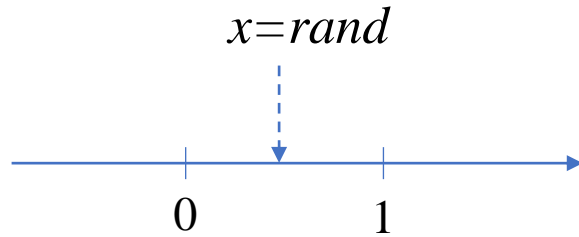
Επιλογή τυχαίας τιμής στο διάστημα  $[a \ b]$



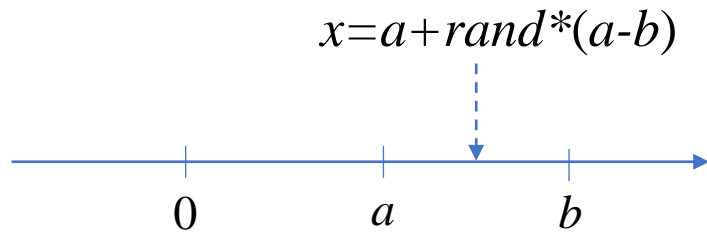
# Αλγόριθμος c-Means

## Τυχαία Επιλογή Αριθμών: Η Εντολή *rand*

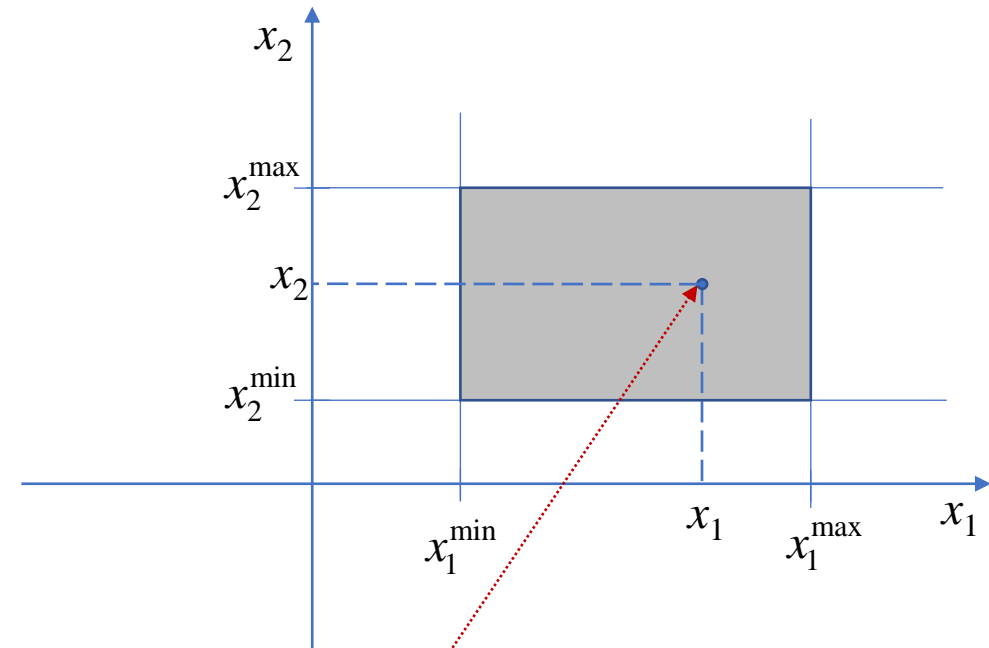
Επιλογή τυχαίας τιμής στο διάστημα  $[0 \ 1]$



Επιλογή τυχαίας τιμής στο διάστημα  $[a \ b]$



Επιλογή τυχαίας τιμής στο παραλληλόγραμμο  $[x_1^{\min} \ x_1^{\max}] \times [x_2^{\min} \ x_2^{\max}]$



$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1^{\min} \\ x_2^{\min} \end{bmatrix} + rand * \begin{bmatrix} x_1^{\max} - x_1^{\min} \\ x_2^{\max} - x_2^{\min} \end{bmatrix}$$

# Αλγόριθμος c-Means

Τυχαία Επιλογή των Κέντρων των Συστάδων: Χρήση της Εντολής *rand*

Πίνακας  $x$

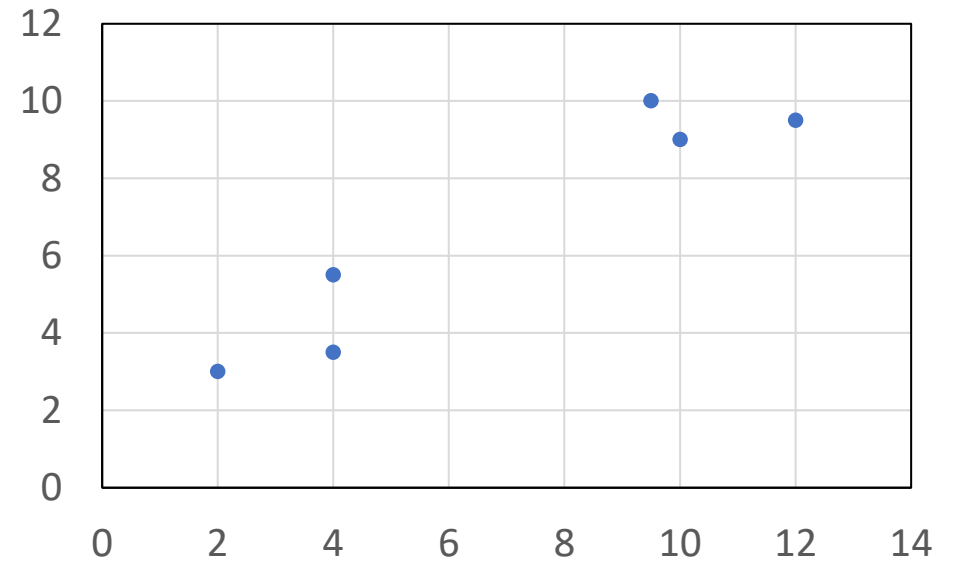
	$x_1$	$x_2$
$x_1$	10	9
$x_2$	4	3.5
$x_3$	9.5	10
$x_4$	2	3
$x_5$	12	9.5
$x_6$	4	5.5

# Αλγόριθμος c-Means

Τυχαία Επιλογή των Κέντρων των Συστάδων: Χρήση της Εντολής *rand*

Πίνακας  $x$

	$x_1$	$x_2$
$x_1$	10	9
$x_2$	4	3.5
$x_3$	9.5	10
$x_4$	2	3
$x_5$	12	9.5
$x_6$	4	5.5

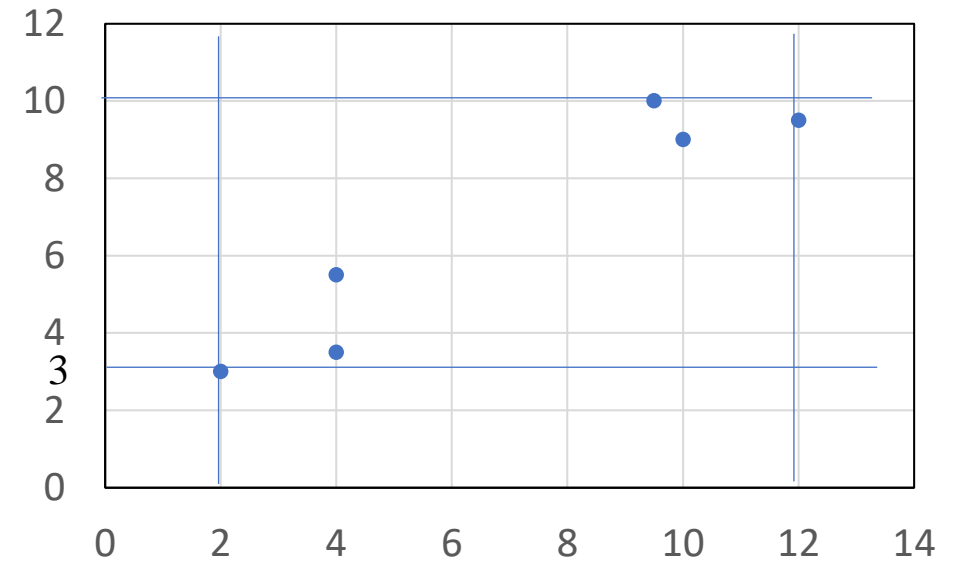


# Αλγόριθμος c-Means

Τυχαία Επιλογή των Κέντρων των Συστάδων: Χρήση της Εντολής *rand*

Πίνακας  $x$

	$x_1$	$x_2$
$x_1$	10	9
$x_2$	4	3.5
$x_3$	9.5	10
$x_4$	2	3
$x_5$	12	9.5
$x_6$	4	5.5



Πεδίο Ορισμού της Ιδιότητας  $x_1$ :  $\text{ΠΟ}_1 = [2 \quad 12]$

Πεδίο Ορισμού της Ιδιότητας  $x_2$ :  $\text{ΠΟ}_2 = [3 \quad 10]$

# Αλγόριθμος c-Means

Τυχαία Επιλογή των Κέντρων των Συστάδων: Χρήση της Εντολής *rand*

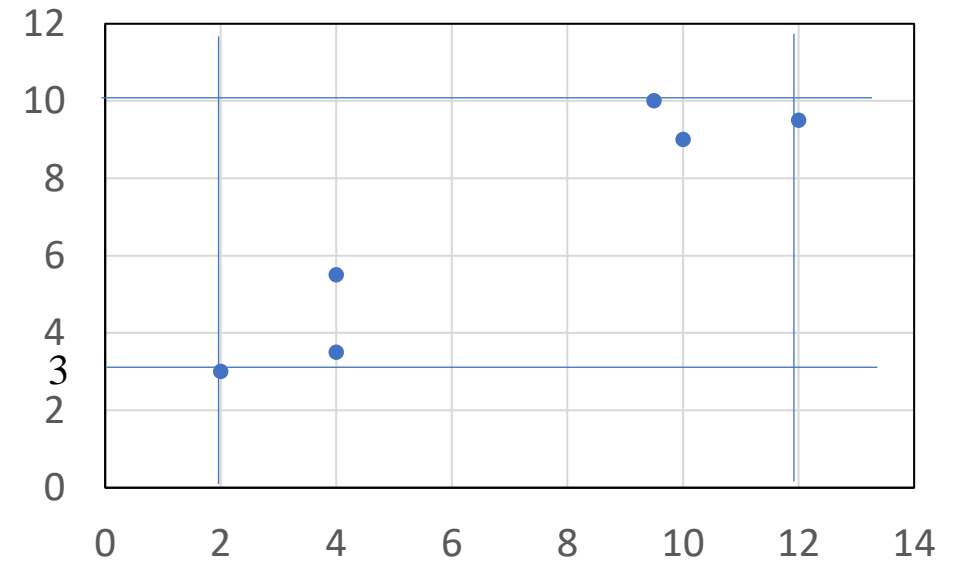
Πίνακας  $x$

	$x_1$	$x_2$
$x_1$	10	9
$x_2$	4	3.5
$x_3$	9.5	10
$x_4$	2	3
$x_5$	12	9.5
$x_6$	4	5.5

Εύρεση ελάχιστου και μέγιστου  
στηλών του πίνακα  $x$

$x_{\min} = \min(x)$ ;

$x_{\max} = \max(x)$ ;



Πεδίο Ορισμού της Ιδιότητας  $x_1$ :  $ΠΟ_1 = [2 \quad 12]$

Πεδίο Ορισμού της Ιδιότητας  $x_2$ :  $ΠΟ_2 = [3 \quad 10]$



# Αλγόριθμος c-Means

Τυχαία Επιλογή των Κέντρων των Συστάδων: Χρήση της Εντολής *rand*

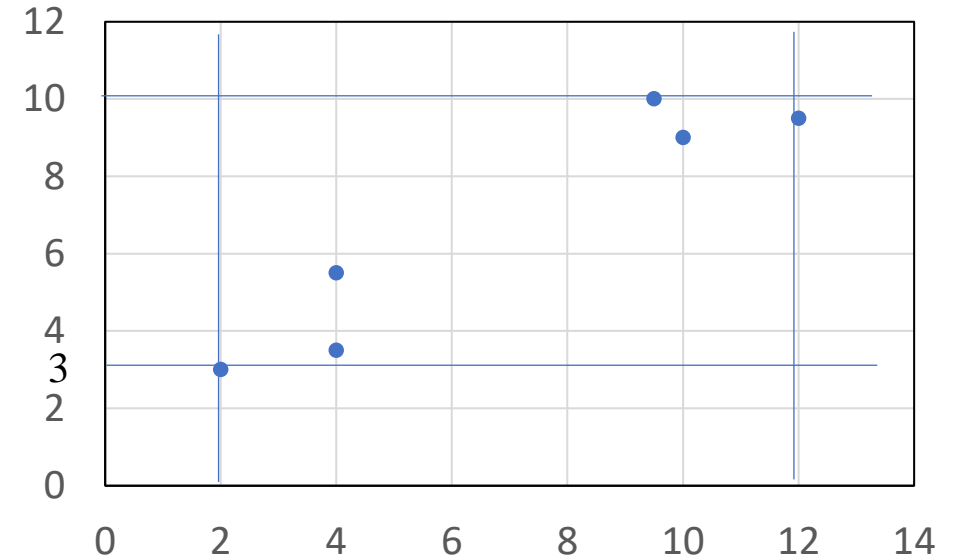
Πίνακας  $x$

	$x_1$	$x_2$
$x_1$	10	9
$x_2$	4	3.5
$x_3$	9.5	10
$x_4$	2	3
$x_5$	12	9.5
$x_6$	4	5.5

Εύρεση ελάχιστου και μέγιστου  
στηλών του πίνακα  $x$

$x_{\min} = \min(x);$   
 $x_{\max} = \max(x);$

$x_{\min} = [2 \quad 3]$   
 $x_{\max} = [12 \quad 10]$



Πεδίο Ορισμού της Ιδιότητας  $x_1$ :  $ΠΟ_1 = [2 \quad 12]$

Πεδίο Ορισμού της Ιδιότητας  $x_2$ :  $ΠΟ_2 = [3 \quad 10]$

# Αλγόριθμος c-Means

Τυχαία Επιλογή των Κέντρων των Συστάδων: Χρήση της Εντολής *rand*

Πίνακας  $x$

	$x_1$	$x_2$
$x_1$	10	9
$x_2$	4	3.5
$x_3$	9.5	10
$x_4$	2	3
$x_5$	12	9.5
$x_6$	4	5.5

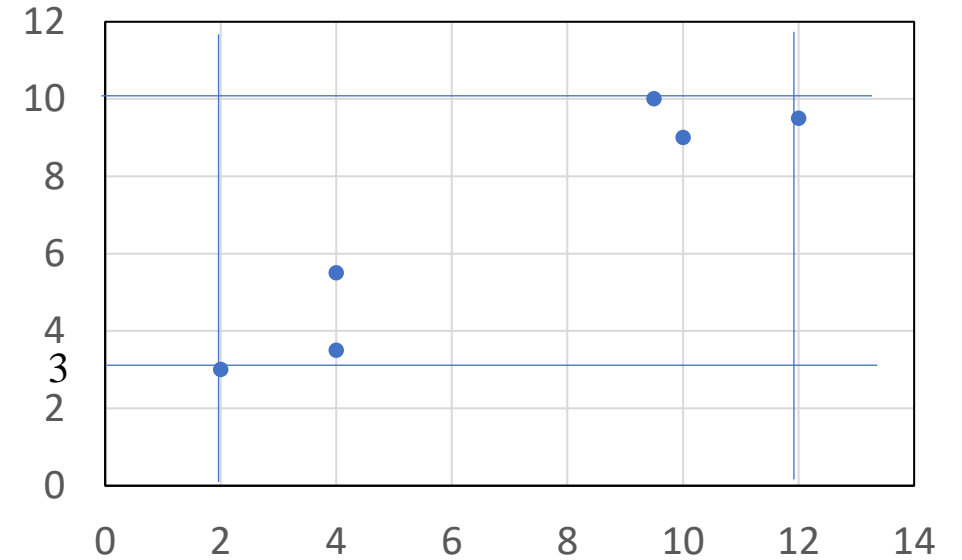
Εύρεση ελάχιστου και μέγιστου  
στηλών του πίνακα  $x$

$x_{\min} = \min(x)$ ;

$x_{\max} = \max(x)$ ;

$x_{\min} = [2 \quad 3]$

$x_{\max} = [12 \quad 10]$



Πεδίο Ορισμού της Ιδιότητας  $x_1$ :  $ΠΟ_1 = [2 \quad 12]$

Πεδίο Ορισμού της Ιδιότητας  $x_2$ :  $ΠΟ_2 = [3 \quad 10]$

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} x_1^{\min} \\ x_2^{\min} \end{bmatrix} + rand * \begin{bmatrix} x_1^{\max} - x_1^{\min} \\ x_2^{\max} - x_2^{\min} \end{bmatrix}$$

# Αλγόριθμος c-Means

Τυχαία Επιλογή των Κέντρων των Συστάδων: Χρήση της Εντολής *rand*

Πίνακας  $x$

	$x_1$	$x_2$
$x_1$	10	9
$x_2$	4	3.5
$x_3$	9.5	10
$x_4$	2	3
$x_5$	12	9.5
$x_6$	4	5.5

Εύρεση ελάχιστου και μέγιστου  
στηλών του πίνακα  $x$

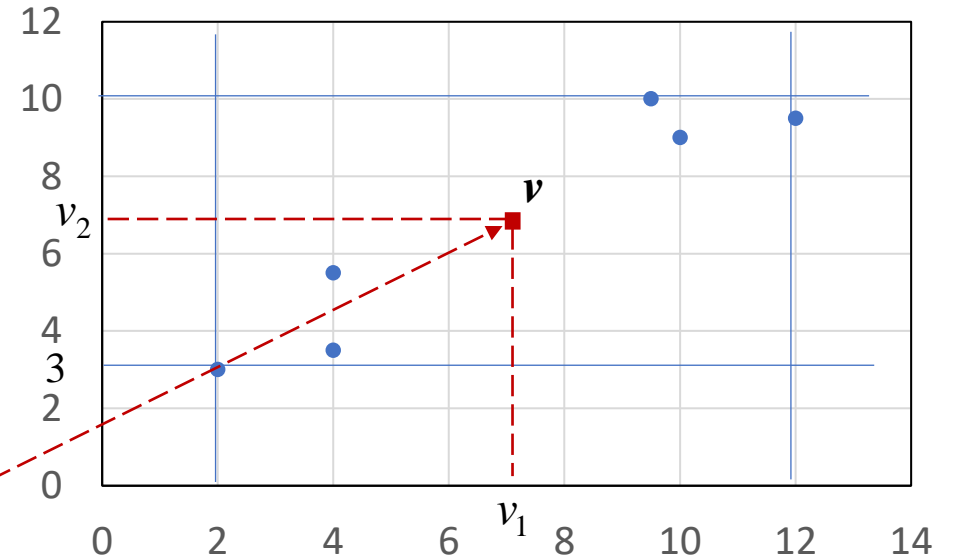
$x_{\min} = \min(x)$ ;

$x_{\max} = \max(x)$ ;

$x_{\min} = [2 \quad 3]$

$x_{\max} = [12 \quad 10]$

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} x_1^{\min} \\ x_2^{\min} \end{bmatrix} + rand * \begin{bmatrix} x_1^{\max} - x_1^{\min} \\ x_2^{\max} - x_2^{\min} \end{bmatrix}$$



Πεδίο Ορισμού της Ιδιότητας  $x_1$ :  $\text{ΠΟ}_1 = [2 \quad 12]$

Πεδίο Ορισμού της Ιδιότητας  $x_2$ :  $\text{ΠΟ}_2 = [3 \quad 10]$

# Αλγόριθμος c-Means

Τυχαία Επιλογή των Κέντρων των Συστάδων: Χρήση της Εντολής *rand*

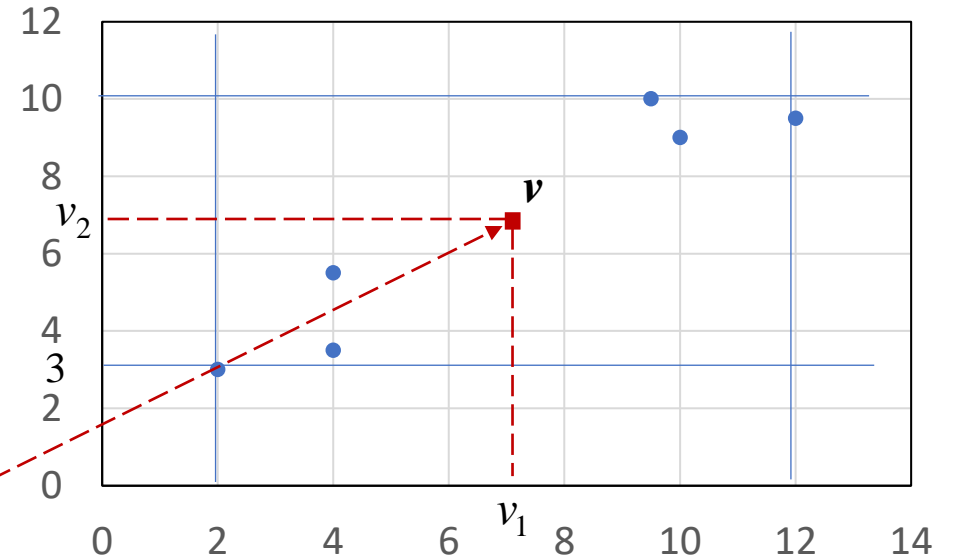
Πίνακας  $x$

	$x_1$	$x_2$
$x_1$	10	9
$x_2$	4	3.5
$x_3$	9.5	10
$x_4$	2	3
$x_5$	12	9.5
$x_6$	4	5.5

Εύρεση ελάχιστου και μέγιστου  
στηλών του πίνακα  $x$

$x_{\min} = \min(x);$   
 $x_{\max} = \max(x);$

$x_{\min} = [2 \quad 3]$   
 $x_{\max} = [12 \quad 10]$



Πεδίο Ορισμού της Ιδιότητας  $x_1$ :  $ΠΟ_1 = [2 \quad 12]$

Πεδίο Ορισμού της Ιδιότητας  $x_2$ :  $ΠΟ_2 = [3 \quad 10]$

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} x_1^{\min} \\ x_2^{\min} \end{bmatrix} + rand * \begin{bmatrix} x_1^{\max} - x_1^{\min} \\ x_2^{\max} - x_2^{\min} \end{bmatrix}$$



$v = x_{\min} + rand * (x_{\max} - x_{\min});$

**Ισχύει για αριθμό ιδιοτήτων μεγαλύτερο του 2**

# Αλγόριθμος c-Means

Τυχαία Επιλογή των Κέντρων των Συστάδων: Χρήση της Εντολής *rand*

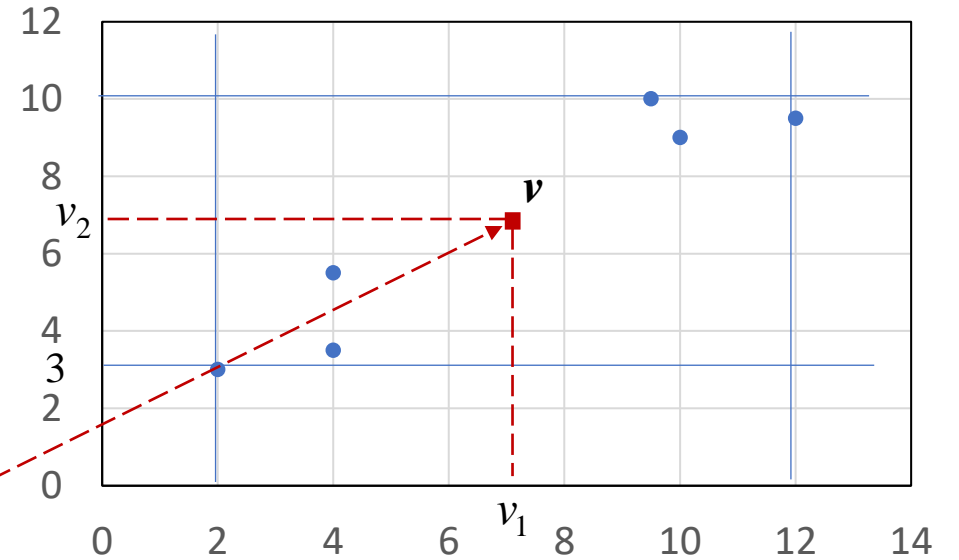
Πίνακας  $x$

	$x_1$	$x_2$
$x_1$	10	9
$x_2$	4	3.5
$x_3$	9.5	10
$x_4$	2	3
$x_5$	12	9.5
$x_6$	4	5.5

Εύρεση ελάχιστου και μέγιστου  
στηλών του πίνακα  $x$

```
x_min=min(x);  
x_max=max(x);  
v=x_min+rand*(x_max-x_min);
```

```
x_min = [2  3]  
x_max = [12 10]
```



Πεδίο Ορισμού της Ιδιότητας  $x_1$ :  $ΠΟ_1 = [2 \ 12]$

Πεδίο Ορισμού της Ιδιότητας  $x_2$ :  $ΠΟ_2 = [3 \ 10]$

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} x_1^{\min} \\ x_2^{\min} \end{bmatrix} + rand * \begin{bmatrix} x_1^{\max} - x_1^{\min} \\ x_2^{\max} - x_2^{\min} \end{bmatrix}$$



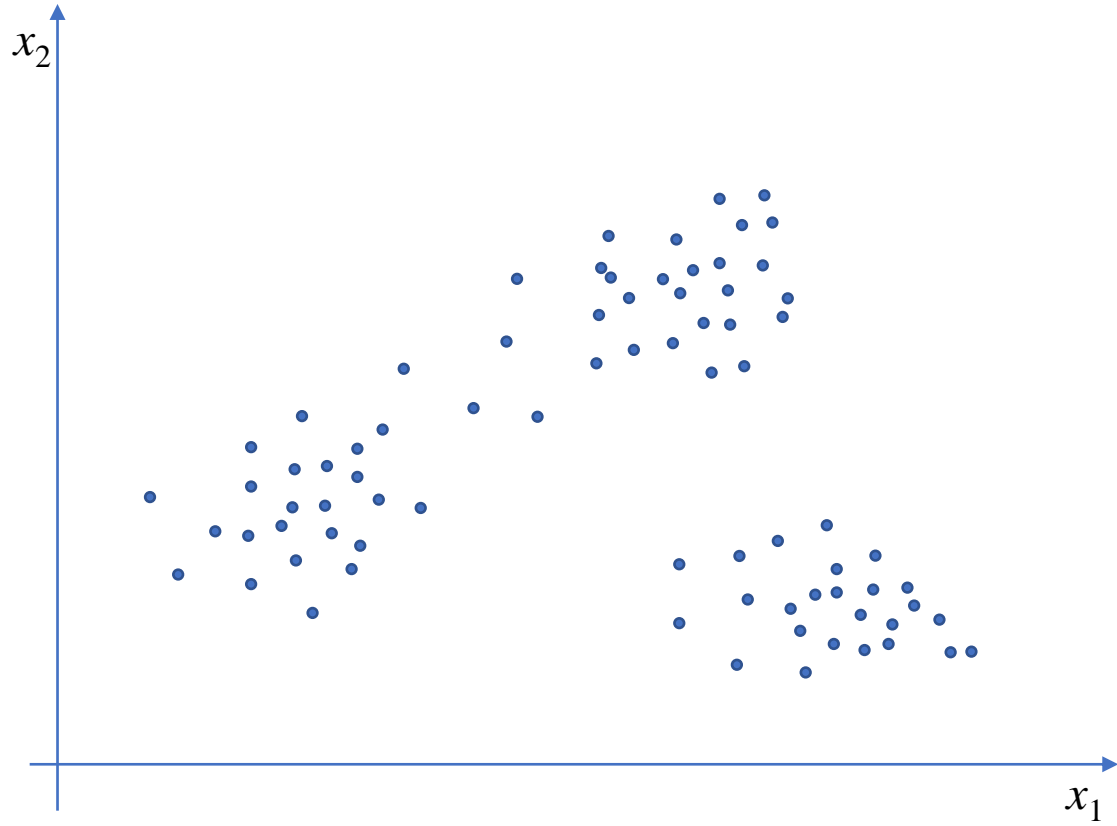
$v = x\_min + rand * (x\_max - x\_min);$

**Ισχύει για αριθμό ιδιοτήτων μεγαλύτερο του 2**

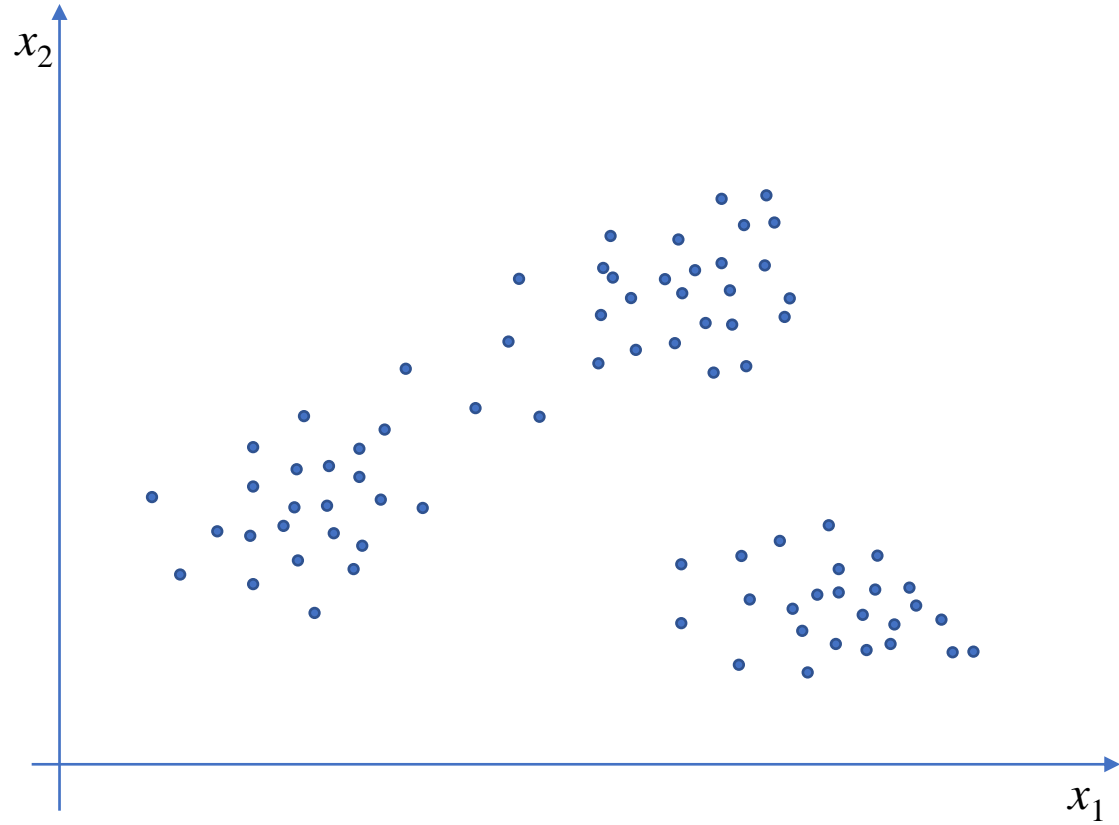
# Αλγόριθμος c-Means

Αρχικοποίηση των παραμέτρων

- Εισάγουμε τον πίνακα των δεδομένων  $x$  και προσδιορίζουμε την διάστασή του  $N \times p$



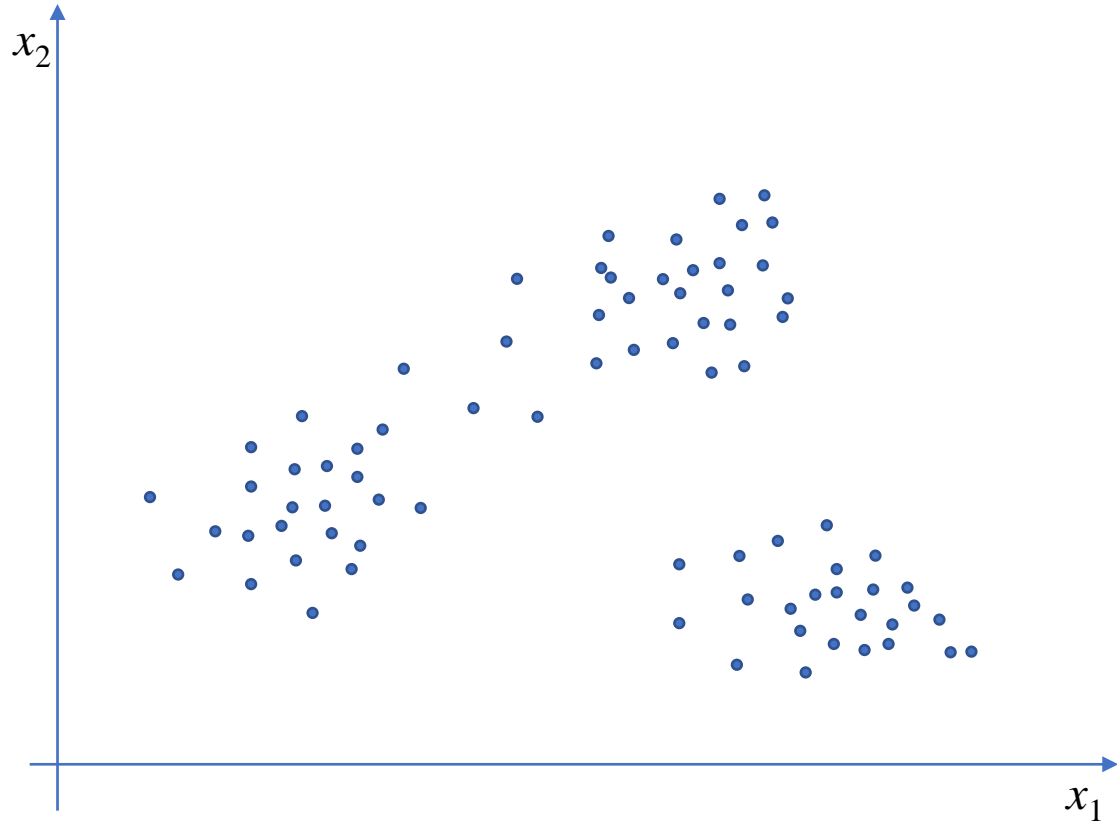
# Αλγόριθμος c-Means



Αρχικοποίηση των παραμέτρων

- Εισάγουμε τον πίνακα των δεδομένων  $x$  και προσδιορίζουμε την διάστασή του  $N \times p$
- Δίνουμε τιμή στον αριθμό των συστάδων  $c$  και μία πολύ μικρή τιμή στην παράμετρο  $\epsilon$

# Αλγόριθμος c-Means

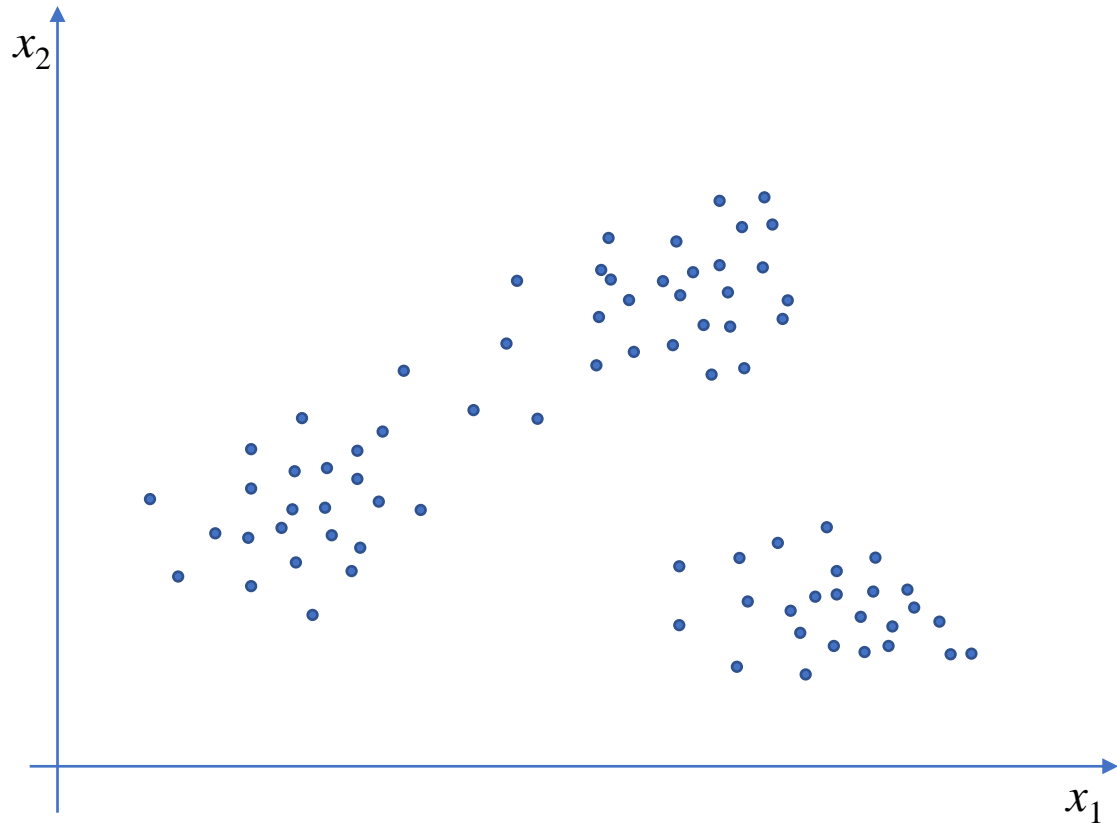


Αρχικοποίηση των παραμέτρων

- Εισάγουμε τον πίνακα των δεδομένων  $x$  και προσδιορίζουμε την διάστασή του  $N \times p$
- Δίνουμε τιμή στον αριθμό των συστάδων  $c$  και μία πολύ μικρή τιμή στην παράμετρο  $\epsilon$
- Δίνουμε τυχαίες τιμές στα κέντρα των συστάδων  $v_1, v_2, \dots, v_c$  που αποθηκεύονται στον πίνακα  $v$ , όπου  $v_1=v(1,:), v_2=v(2,:), \dots, v_c=v(c,:)$ . Άρα ο πίνακας  $v$  είναι  $c \times p$



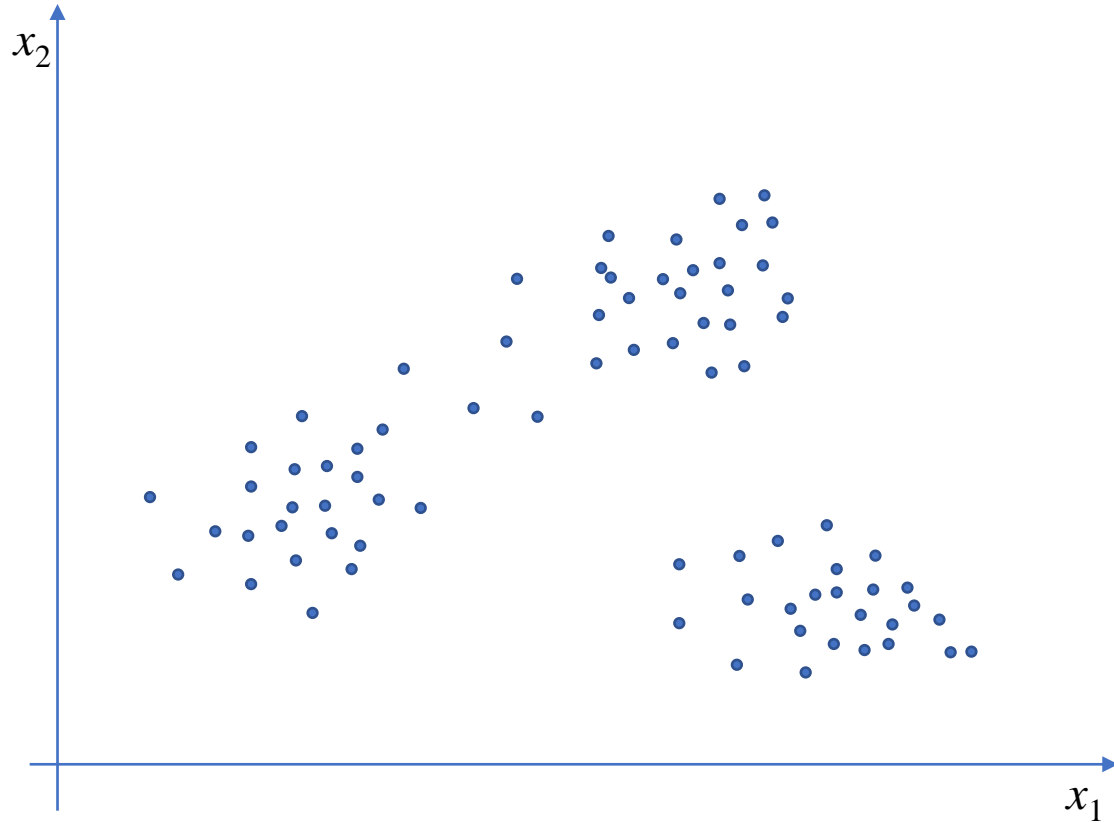
# Αλγόριθμος c-Means



Αρχικοποίηση των παραμέτρων

- Εισάγουμε τον πίνακα των δεδομένων  $x$  και προσδιορίζουμε την διάστασή του  $N \times p$
- Δίνουμε τιμή στον αριθμό των συστάδων  $c$  και μία πολύ μικρή τιμή στην παράμετρο  $\epsilon$
- Δίνουμε τυχαίες τιμές στα κέντρα των συστάδων  $v_1, v_2, \dots, v_c$  που αποθηκεύονται στον πίνακα  $v$ , όπου  $v_1=v(1,:)$ ,  $v_2=v(2,:)$ ,  $\dots$ ,  $v_c=v(c,:)$ . Άρα ο πίνακας  $v$  είναι  $c \times p$
- Δίνουμε μία τιμή στην παράμετρο  $E$  μεγαλύτερη από την τιμή του  $\epsilon$  (δηλαδή  $E > \epsilon$ )

# Αλγόριθμος c-Means



Αρχικοποίηση των παραμέτρων

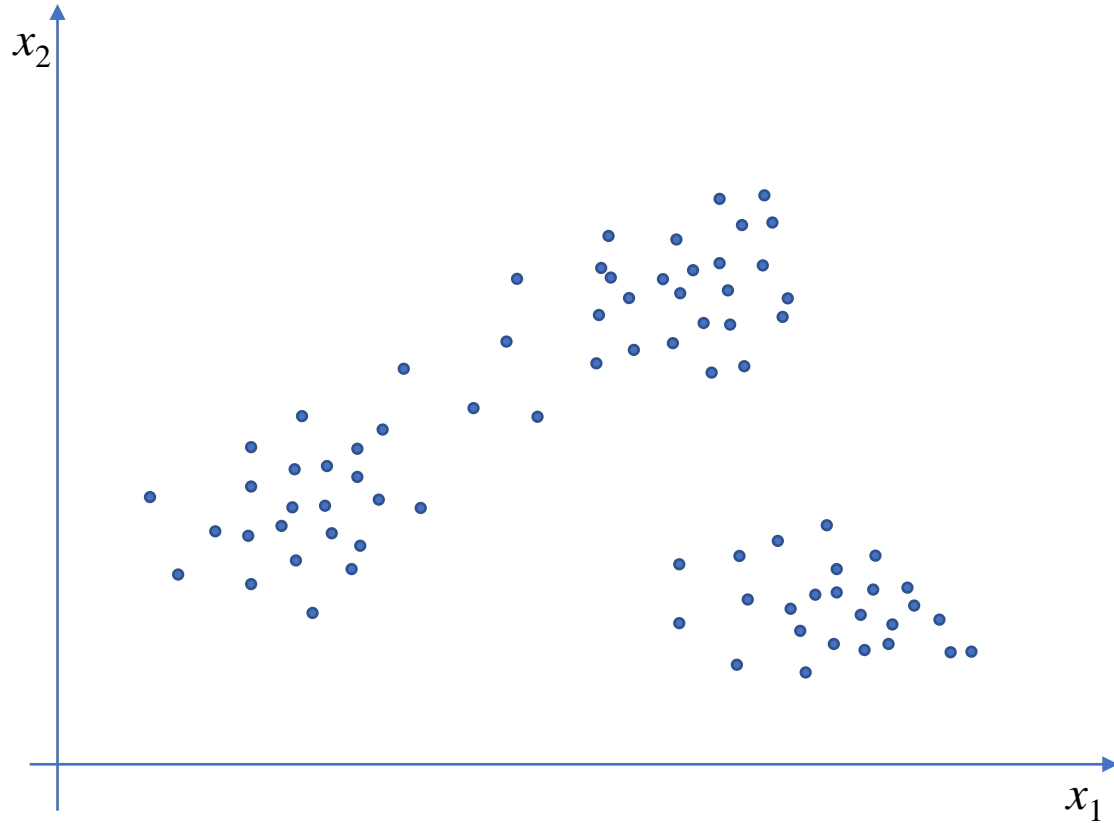
- Εισάγουμε τον πίνακα των δεδομένων  $x$  και προσδιορίζουμε την διάστασή του  $N \times p$
- Δίνουμε τιμή στον αριθμό των συστάδων  $c$  και μία πολύ μικρή τιμή στην παράμετρο  $\epsilon$
- Δίνουμε τυχαίες τιμές στα κέντρα των συστάδων  $v_1, v_2, \dots, v_c$  που αποθηκεύονται στον πίνακα  $v$ , όπου  $v_1=v(1,:)$ ,  $v_2=v(2,:)$ ,  $\dots$ ,  $v_c=v(c,:)$ . Άρα ο πίνακας  $v$  είναι  $c \times p$
- Δίνουμε μία τιμή στην παράμετρο  $E$  μεγαλύτερη από την τιμή του  $\epsilon$  (δηλαδή  $E > \epsilon$ )

Iteration\_Number=0

**While**  $E > \epsilon$  **do**

    Iteration\_Number= Iteration\_Number+1

# Αλγόριθμος c-Means



Αρχικοποίηση των παραμέτρων

- Εισάγουμε τον πίνακα των δεδομένων  $x$  και προσδιορίζουμε την διάστασή του  $N \times p$
- Δίνουμε τιμή στον αριθμό των συστάδων  $c$  και μία πολύ μικρή τιμή στην παράμετρο  $\epsilon$
- Δίνουμε τυχαίες τιμές στα κέντρα των συστάδων  $v_1, v_2, \dots, v_c$  που αποθηκεύονται στον πίνακα  $v$ , όπου  $v_1=v(1,:)$ ,  $v_2=v(2,:)$ ,  $\dots$ ,  $v_c=v(c,:)$ . Άρα ο πίνακας  $v$  είναι  $c \times p$
- Δίνουμε μία τιμή στην παράμετρο  $E$  μεγαλύτερη από την τιμή του  $\epsilon$  (δηλαδή  $E > \epsilon$ )

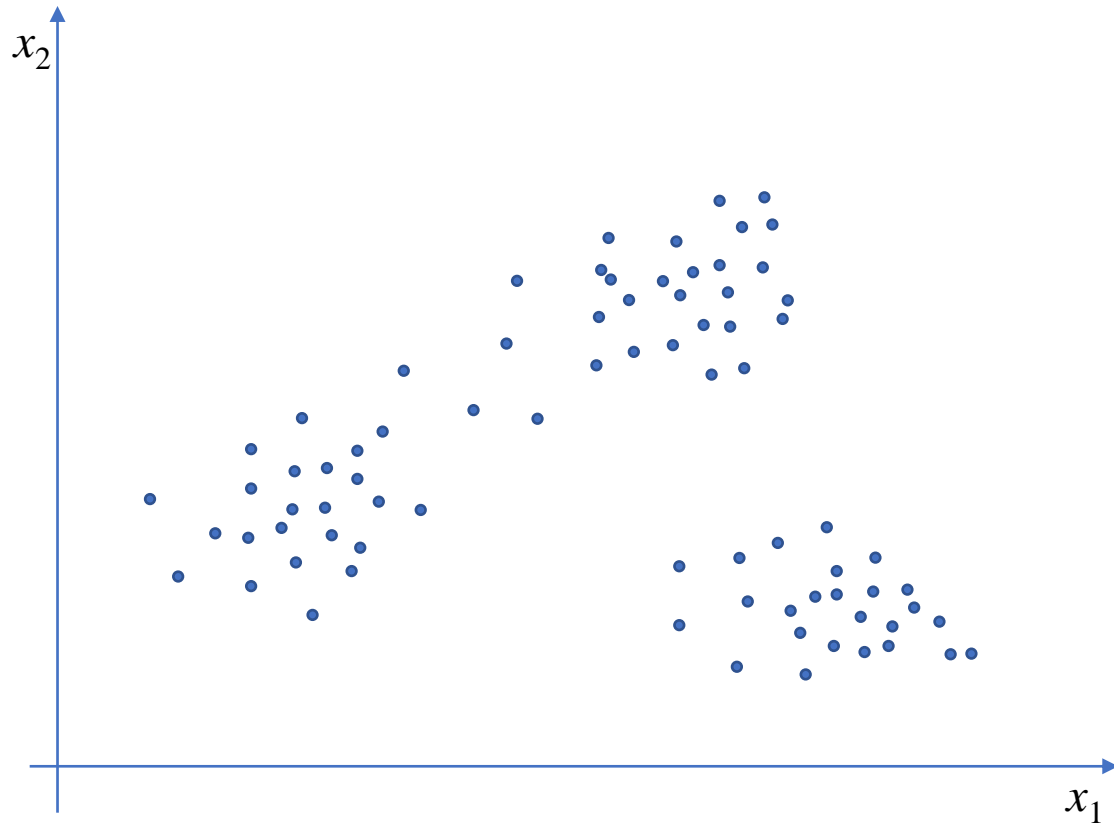
Iteration\_Number=0

**While**  $E > \epsilon$  **do**

Iteration\_Number= Iteration\_Number+1

**Βήμα 1.** Αποθήκευσε τις τωρινές τιμές των κέντρων των συστάδων  $v$  σε έναν πίνακα  $v\_old$  που είναι βοηθητικός για να μην χαθούν αυτές οι τιμές των κέντρων:  $v\_old=v$

# Αλγόριθμος c-Means



Αρχικοποίηση των παραμέτρων

- Εισάγουμε τον πίνακα των δεδομένων  $x$  και προσδιορίζουμε την διάστασή του  $N \times p$
- Δίνουμε τιμή στον αριθμό των συστάδων  $c$  και μία πολύ μικρή τιμή στην παράμετρο  $\epsilon$
- Δίνουμε τυχαίες τιμές στα κέντρα των συστάδων  $v_1, v_2, \dots, v_c$  που αποθηκεύονται στον πίνακα  $v$ , όπου  $v_1=v(1,:)$ ,  $v_2=v(2,:)$ ,  $\dots$ ,  $v_c=v(c,:)$ . Άρα ο πίνακας  $v$  είναι  $c \times p$
- Δίνουμε μία τιμή στην παράμετρο  $E$  μεγαλύτερη από την τιμή του  $\epsilon$  (δηλαδή  $E > \epsilon$ )

Iteration\_Number=0

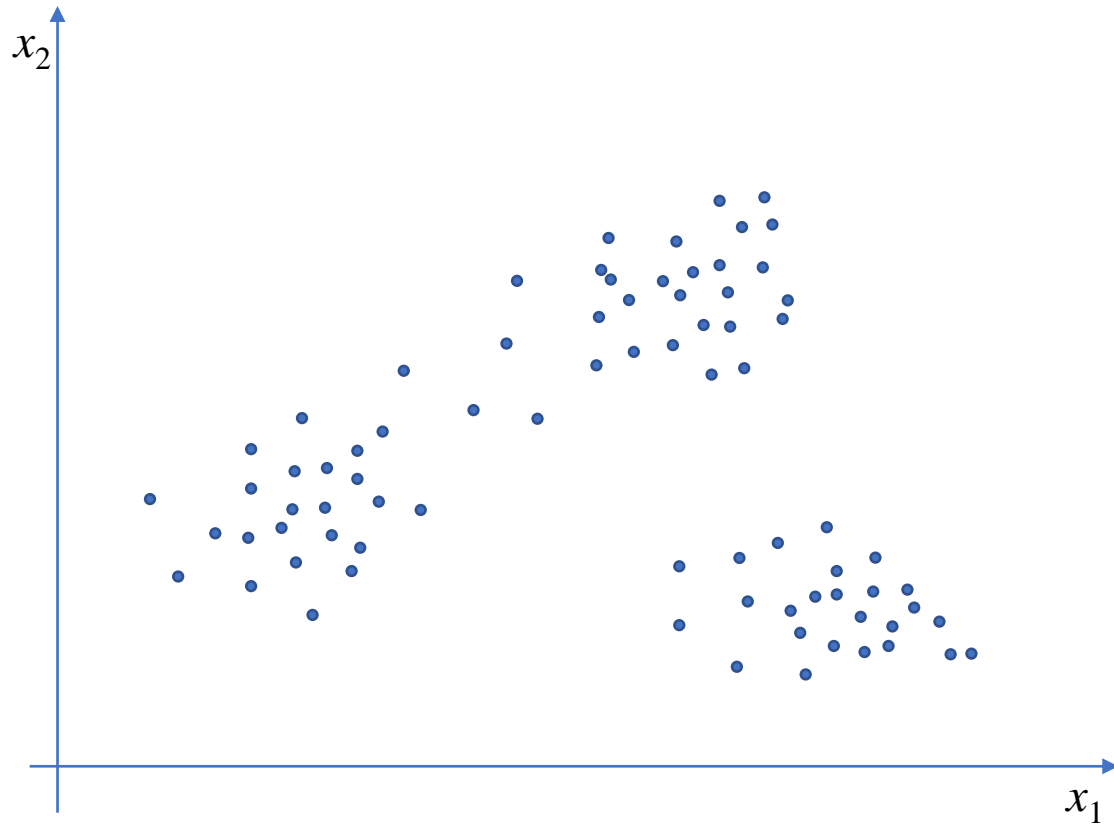
**While**  $E > \epsilon$  **do**

Iteration\_Number= Iteration\_Number+1

**Βήμα 1.** Αποθήκευσε τις τωρινές τιμές των κέντρων των συστάδων  $v$  σε έναν πίνακα  $v\_old$  που είναι βοηθητικός για να μην χαθούν αυτές οι τιμές των κέντρων:  $v\_old=v$

**Βήμα 2.** Δεδομένου ότι τα κέντρα των  $c$  συστάδων που είναι αποθηκευμένα στον πίνακα  $v$  είναι γνωστά υπολόγισε τον πίνακα των συναρτήσεων συμμετοχής  $u$  που έχουν τα δεδομένα στις  $c$  συστάδες

# Αλγόριθμος c-Means



Αρχικοποίηση των παραμέτρων

- Εισάγουμε τον πίνακα των δεδομένων  $x$  και προσδιορίζουμε την διάστασή του  $N \times p$
- Δίνουμε τιμή στον αριθμό των συστάδων  $c$  και μία πολύ μικρή τιμή στην παράμετρο  $\epsilon$
- Δίνουμε τυχαίες τιμές στα κέντρα των συστάδων  $v_1, v_2, \dots, v_c$  που αποθηκεύονται στον πίνακα  $v$ , όπου  $v_1=v(1,:)$ ,  $v_2=v(2,:)$ ,  $\dots$ ,  $v_c=v(c,:)$ . Άρα ο πίνακας  $v$  είναι  $c \times p$
- Δίνουμε μία τιμή στην παράμετρο  $E$  μεγαλύτερη από την τιμή του  $\epsilon$  (δηλαδή  $E > \epsilon$ )

Iteration\_Number=0

**While**  $E > \epsilon$  **do**

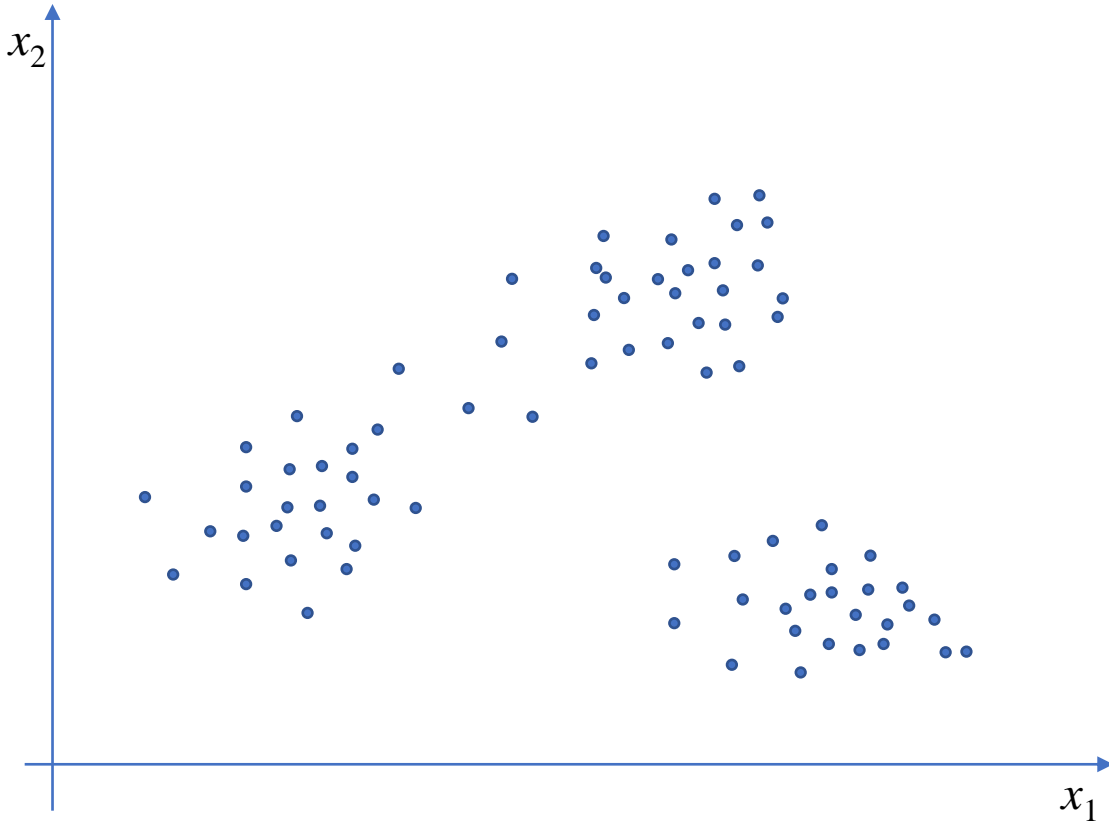
Iteration\_Number= Iteration\_Number+1

**Βήμα 1.** Αποθήκευσε τις τωρινές τιμές των κέντρων των συστάδων  $v$  σε έναν πίνακα  $v\_old$  που είναι βοηθτικός για να μην χαθούν αυτές οι τιμές των κέντρων:  $v\_old=v$

**Βήμα 2.** Δεδομένου ότι τα κέντρα των  $c$  συστάδων που είναι αποθηκευμένα στον πίνακα  $v$  είναι γνωστά υπολόγισε τον πίνακα των συναρτήσεων συμμετοχής  $u$  που έχουν τα δεδομένα στις  $c$  συστάδες

**Βήμα 3.** Δεδομένου ότι οι συναρτήσεις συμμετοχής των δεδομένων στις συστάδες, που είναι αποθηκευμένες στον πίνακα  $u$  είναι γνωστές, υπολόγισε τις καινούργιες τιμές των κέντρων  $v$  (δηλαδή οι τιμές του πίνακα  $v$  έχουν αλλάξει)

# Αλγόριθμος c-Means



Αρχικοποίηση των παραμέτρων

- Εισάγουμε τον πίνακα των δεδομένων  $x$  και προσδιορίζουμε την διάστασή του  $N \times p$
- Δίνουμε τιμή στον αριθμό των συστάδων  $c$  και μία πολύ μικρή τιμή στην παράμετρο  $\epsilon$
- Δίνουμε τυχαίες τιμές στα κέντρα των συστάδων  $v_1, v_2, \dots, v_c$  που αποθηκεύονται στον πίνακα  $v$ , όπου  $v_1=v(1,:)$ ,  $v_2=v(2,:)$ ,  $\dots$ ,  $v_c=v(c,:)$ . Άρα ο πίνακας  $v$  είναι  $c \times p$
- Δίνουμε μία τιμή στην παράμετρο  $E$  μεγαλύτερη από την τιμή του  $\epsilon$  (δηλαδή  $E > \epsilon$ )

Iteration\_Number=0

**While**  $E > \epsilon$  **do**

Iteration\_Number= Iteration\_Number+1

**Βήμα 1.** Αποθήκευσε τις τωρινές τιμές των κέντρων των συστάδων  $v$  σε έναν πίνακα  $v\_old$  που είναι βοηθητικός για να μην χαθούν αυτές οι τιμές των κέντρων:  $v\_old=v$

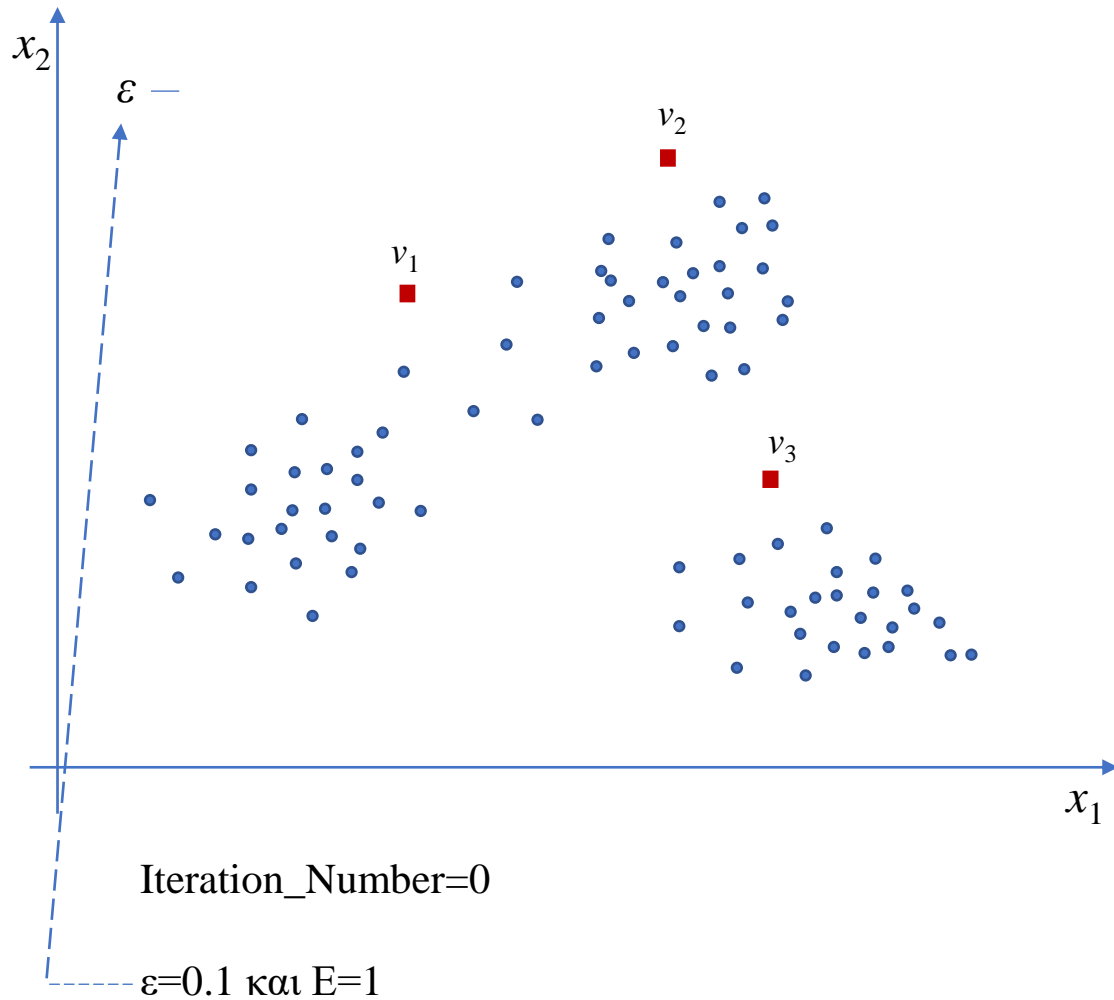
**Βήμα 2.** Δεδομένου ότι τα κέντρα των  $c$  συστάδων που είναι αποθηκευμένα στον πίνακα  $v$  είναι γνωστά υπολόγισε τον πίνακα των συναρτήσεων συμμετοχής  $u$  που έχουν τα δεδομένα στις  $c$  συστάδες

**Βήμα 3.** Δεδομένου ότι οι συναρτήσεις συμμετοχής των δεδομένων στις συστάδες, που είναι αποθηκευμένες στον πίνακα  $u$  είναι γνωστές, υπολόγισε τις καινούργιες τιμές των κέντρων  $v$  (δηλαδή οι τιμές του πίνακα  $v$  έχουν αλλάξει)

**Βήμα 4.** Για κάθε  $i=1, 2, \dots, c$  υπολογίζουμε την απόσταση  $d(i)=\sqrt{(v(i,1)-v\_old(i,1))^2+(v(i,2)-v\_old(i,2))^2+\dots+(v(i,p)-v\_old(i,p))^2}$  και υπολογίζουμε την μέγιστη από όλες αυτές:  
 $E=\max\{d(1), d(2), \dots, d(c)\}$

**EndWhile**

# Αλγόριθμος c-Means



Στο σχήμα φαίνονται οι τυχαίες θέσεις των κέντρων και το μήκος (τιμή) του  $\epsilon$

Αρχικοποίηση των παραμέτρων

- Εισάγουμε τον πίνακα των δεδομένων  $x$  και προσδιορίζουμε την διάστασή του  $N \times p$
- Δίνουμε τιμή στον αριθμό των συστάδων  $c$  και μία πολύ μικρή τιμή στην παράμετρο  $\epsilon$
- Δίνουμε τυχαίες τιμές στα κέντρα των συστάδων  $v_1, v_2, \dots, v_c$  που αποθηκεύονται στον πίνακα  $v$ , όπου  $v_1=v(1,:)$ ,  $v_2=v(2,:)$ ,  $\dots$ ,  $v_c=v(c,:)$ . Άρα ο πίνακας  $v$  είναι  $c \times p$
- Δίνουμε μία τιμή στην παράμετρο  $E$  μεγαλύτερη από την τιμή του  $\epsilon$  (δηλαδή  $E > \epsilon$ )

Iteration\_Number=0

**While**  $E > \epsilon$  **do**

Iteration\_Number= Iteration\_Number+1

**Βήμα 1.** Αποθήκευσε τις τωρινές τιμές των κέντρων των συστάδων  $v$  σε έναν πίνακα  $v\_old$  που είναι βοηθητικός για να μην χαθούν αυτές οι τιμές των κέντρων:  $v\_old=v$

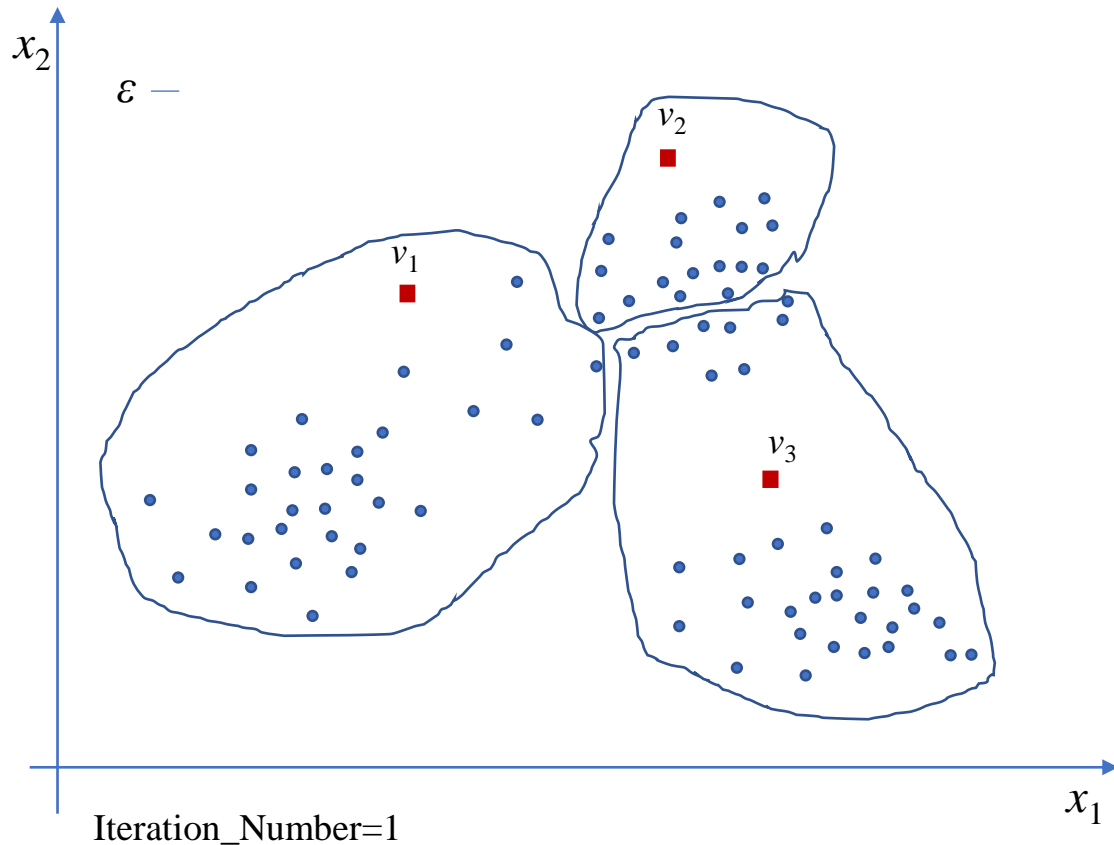
**Βήμα 2.** Δεδομένου ότι τα κέντρα των  $c$  συστάδων που είναι αποθηκευμένα στον πίνακα  $v$  είναι γνωστά υπολόγισε τον πίνακα των συναρτήσεων συμμετοχής  $u$  που έχουν τα δεδομένα στις  $c$  συστάδες

**Βήμα 3.** Δεδομένου ότι οι συναρτήσεις συμμετοχής των δεδομένων στις συστάδες, που είναι αποθηκευμένες στον πίνακα  $u$  είναι γνωστές, υπολόγισε τις καινούργιες τιμές των κέντρων  $v$  (δηλαδή οι τιμές του πίνακα  $v$  έχουν αλλάξει)

**Βήμα 4.** Για κάθε  $i=1, 2, \dots, c$  υπολογίζουμε την απόσταση  $d(i)=\text{sqrt}((v(i,1)-v\_old(i,1))^2+(v(i,2)-v\_old(i,2))^2+\dots+(v(i,p)-v\_old(i,p))^2)$  και υπολογίζουμε την μέγιστη από όλες αυτές:  
 $E=\max\{d(1), d(2), \dots, d(c)\}$

**EndWhile**

# Αλγόριθμος c-Means



Βήμα 2. Δίνει τις συστάδες του σχήματος

Αρχικοποίηση των παραμέτρων

- Εισάγουμε τον πίνακα των δεδομένων  $x$  και προσδιορίζουμε την διάστασή του  $N \times p$
- Δίνουμε τιμή στον αριθμό των συστάδων  $c$  και μία πολύ μικρή τιμή στην παράμετρο  $\epsilon$
- Δίνουμε τυχαίες τιμές στα κέντρα των συστάδων  $v_1, v_2, \dots, v_c$  που αποθηκεύονται στον πίνακα  $v$ , όπου  $v_1=v(1,:)$ ,  $v_2=v(2,:)$ ,  $\dots$ ,  $v_c=v(c,:)$ . Άρα ο πίνακας  $v$  είναι  $c \times p$
- Δίνουμε μία τιμή στην παράμετρο  $E$  μεγαλύτερη από την τιμή του  $\epsilon$  (δηλαδή  $E > \epsilon$ )

Iteration\_Number=0

**While**  $E > \epsilon$  **do**

Iteration\_Number= Iteration\_Number+1

**Βήμα 1.** Αποθήκευσε τις τωρινές τιμές των κέντρων των συστάδων  $v$  σε έναν πίνακα  $v\_old$  που είναι βοηθητικός για να μην χαθούν αυτές οι τιμές των κέντρων:  $v\_old=v$

**Βήμα 2.** Δεδομένου ότι τα κέντρα των  $c$  συστάδων που είναι αποθηκευμένα στον πίνακα  $v$  είναι γνωστά υπολόγισε τον πίνακα των συναρτήσεων συμμετοχής  $u$  που έχουν τα δεδομένα στις  $c$  συστάδες

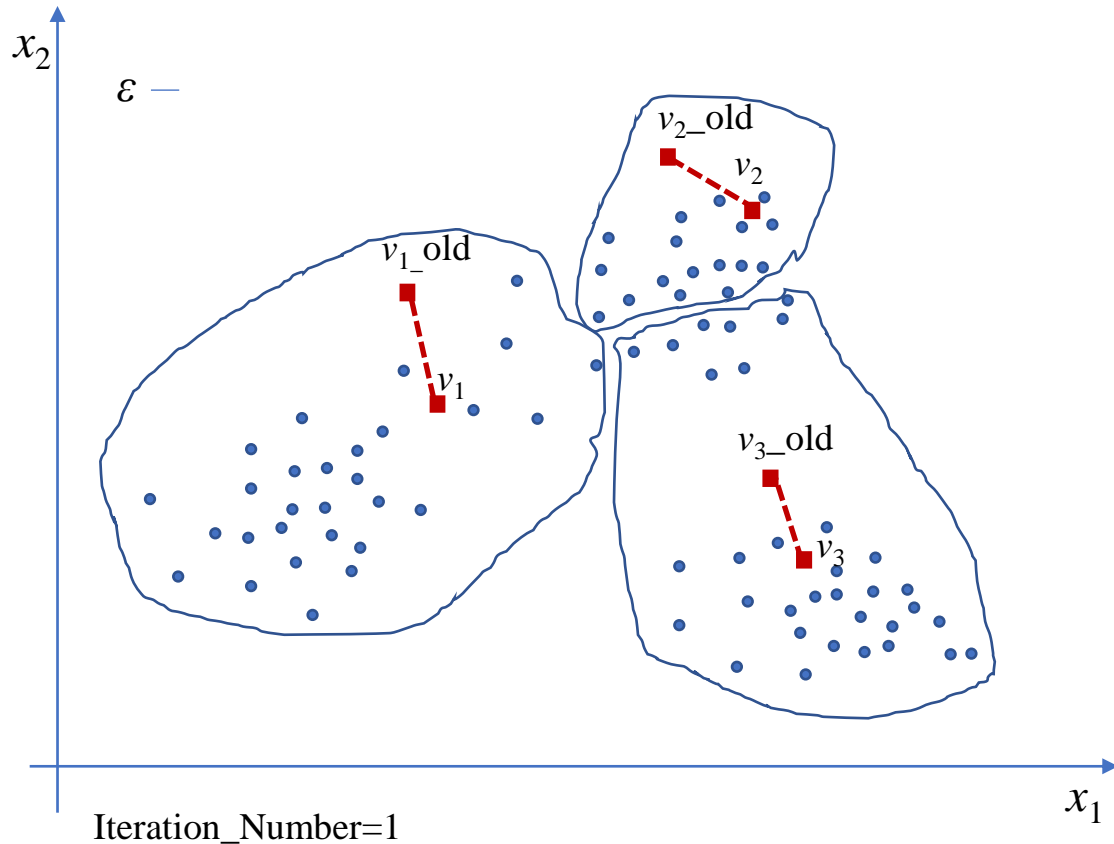
**Βήμα 3.** Δεδομένου ότι οι συναρτήσεις συμμετοχής των δεδομένων στις συστάδες, που είναι αποθηκευμένες στον πίνακα  $u$  είναι γνωστές, υπολόγισε τις καινούργιες τιμές των κέντρων  $v$  (δηλαδή οι τιμές του πίνακα  $v$  έχουν αλλάξει)

**Βήμα 4.** Για κάθε  $i=1, 2, \dots, c$  υπολογίζουμε την απόσταση  $d(i)=\sqrt{(v(i,1)-v\_old(i,1))^2+(v(i,2)-v\_old(i,2))^2+\dots+(v(i,p)-v\_old(i,p))^2}$  και υπολογίζουμε την μέγιστη από όλες αυτές:  $E=\max\{d(1), d(2), \dots, d(c)\}$

**EndWhile**



# Αλγόριθμος c-Means



Iteration\_Number=1

Βήμα 2. Δίνει τις συστάδες του σχήματος

Βήμα 3. Τα κέντρα μετακινούνται σε καινούργιες τιμές με κατεύθυνση προς τον κύριο όγκο δεδομένων σε κάθε συστάδα

Αρχικοποίηση των παραμέτρων

- Εισάγουμε τον πίνακα των δεδομένων  $x$  και προσδιορίζουμε την διάστασή του  $N \times p$
- Δίνουμε τιμή στον αριθμό των συστάδων  $c$  και μία πολύ μικρή τιμή στην παράμετρο  $\epsilon$
- Δίνουμε τυχαίες τιμές στα κέντρα των συστάδων  $v_1, v_2, \dots, v_c$  που αποθηκεύονται στον πίνακα  $v$ , όπου  $v_1=v(1,:)$ ,  $v_2=v(2,:)$ ,  $\dots$ ,  $v_c=v(c,:)$ . Άρα ο πίνακας  $v$  είναι  $c \times p$
- Δίνουμε μία τιμή στην παράμετρο  $E$  μεγαλύτερη από την τιμή του  $\epsilon$  (δηλαδή  $E > \epsilon$ )

Iteration\_Number=0

**While**  $E > \epsilon$  **do**

Iteration\_Number= Iteration\_Number+1

**Βήμα 1.** Αποθήκευσε τις τωρινές τιμές των κέντρων των συστάδων  $v$  σε έναν πίνακα  $v\_old$  που είναι βοηθητικός για να μην χαθούν αυτές οι τιμές των κέντρων:  $v\_old=v$

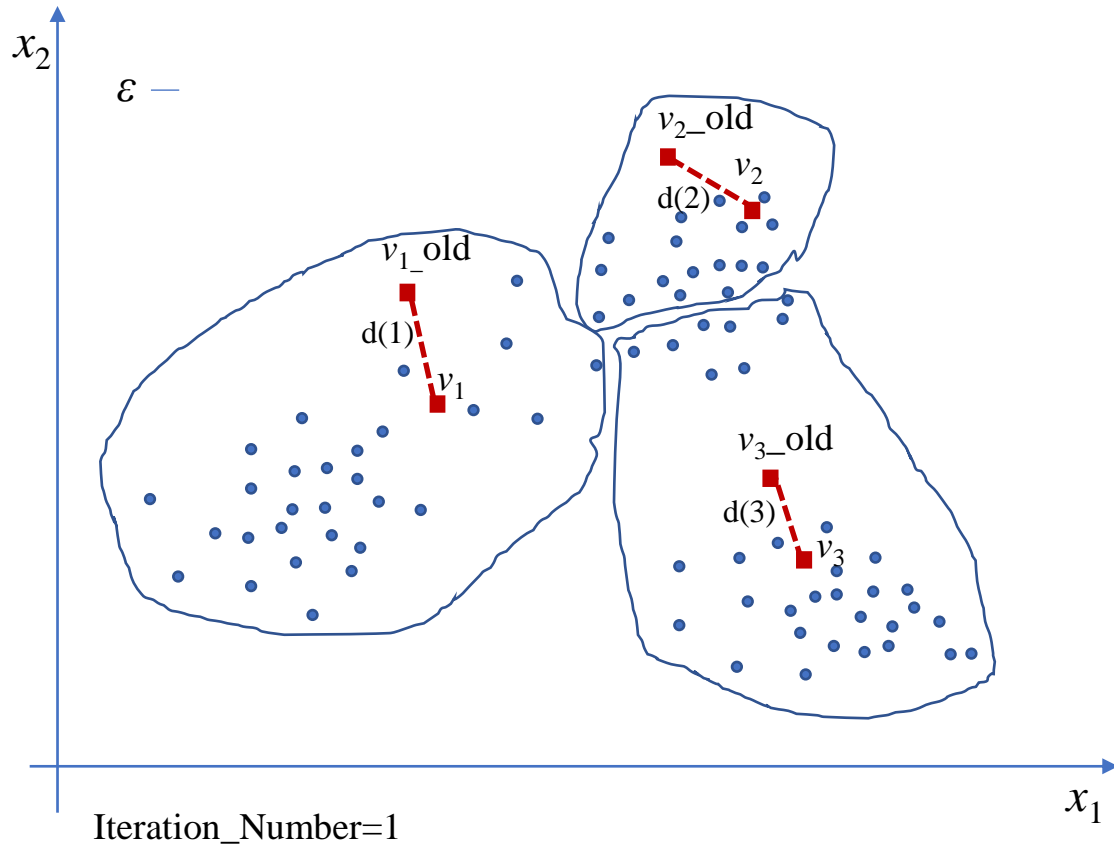
**Βήμα 2.** Δεδομένου ότι τα κέντρα των  $c$  συστάδων που είναι αποθηκευμένα στον πίνακα  $v$  είναι γνωστά υπολόγισε τον πίνακα των συναρτήσεων συμμετοχής  $u$  που έχουν τα δεδομένα στις  $c$  συστάδες

**Βήμα 3.** Δεδομένου ότι οι συναρτήσεις συμμετοχής των δεδομένων στις συστάδες, που είναι αποθηκευμένες στον πίνακα  $u$  είναι γνωστές, υπολόγισε τις καινούργιες τιμές των κέντρων  $v$  (δηλαδή οι τιμές του πίνακα  $v$  έχουν αλλάξει)

**Βήμα 4.** Για κάθε  $i=1, 2, \dots, c$  υπολογίζουμε την απόσταση  $d(i)=\sqrt{(v(i,1)-v\_old(i,1))^2+(v(i,2)-v\_old(i,2))^2+\dots+(v(i,p)-v\_old(i,p))^2}$  και υπολογίζουμε την μέγιστη από όλες αυτές:  $E=\max\{d(1), d(2), \dots, d(c)\}$

**EndWhile**

# Αλγόριθμος c-Means



Iteration\_Number=1

Βήμα 2. Δίνει τις συστάδες του σχήματος

Βήμα 3. Τα κέντρα μετακινούνται σε καινούργιες τιμές με κατεύθυνση προς τον κύριο όγκο δεδομένων σε κάθε συστάδα

Βήμα 4. Στο σχήμα φαίνονται οι αποστάσεις  $d(1)$ ,  $d(2)$  και  $d(3)$  και είναι προφανές ότι  $E=d(1)$ , το οποίο είναι  $E>\epsilon$

Αρχικοποίηση των παραμέτρων

- Εισάγουμε τον πίνακα των δεδομένων  $x$  και προσδιορίζουμε την διάστασή του  $N \times p$
- Δίνουμε τιμή στον αριθμό των συστάδων  $c$  και μία πολύ μικρή τιμή στην παράμετρο  $\epsilon$
- Δίνουμε τυχαίες τιμές στα κέντρα των συστάδων  $v_1, v_2, \dots, v_c$  που αποθηκεύονται στον πίνακα  $v$ , όπου  $v_1=v(1,:)$ ,  $v_2=v(2,:)$ ,  $\dots$ ,  $v_c=v(c,:)$ . Άρα ο πίνακας  $v$  είναι  $c \times p$
- Δίνουμε μία τιμή στην παράμετρο  $E$  μεγαλύτερη από την τιμή του  $\epsilon$  (δηλαδή  $E>\epsilon$ )

Iteration\_Number=0

**While**  $E>\epsilon$  **do**

Iteration\_Number=Iteration\_Number+1

**Βήμα 1.** Αποθήκευσε τις τωρινές τιμές των κέντρων των συστάδων  $v$  σε έναν πίνακα  $v\_old$  που είναι βοηθητικός για να μην χαθούν αυτές οι τιμές των κέντρων:  $v\_old=v$

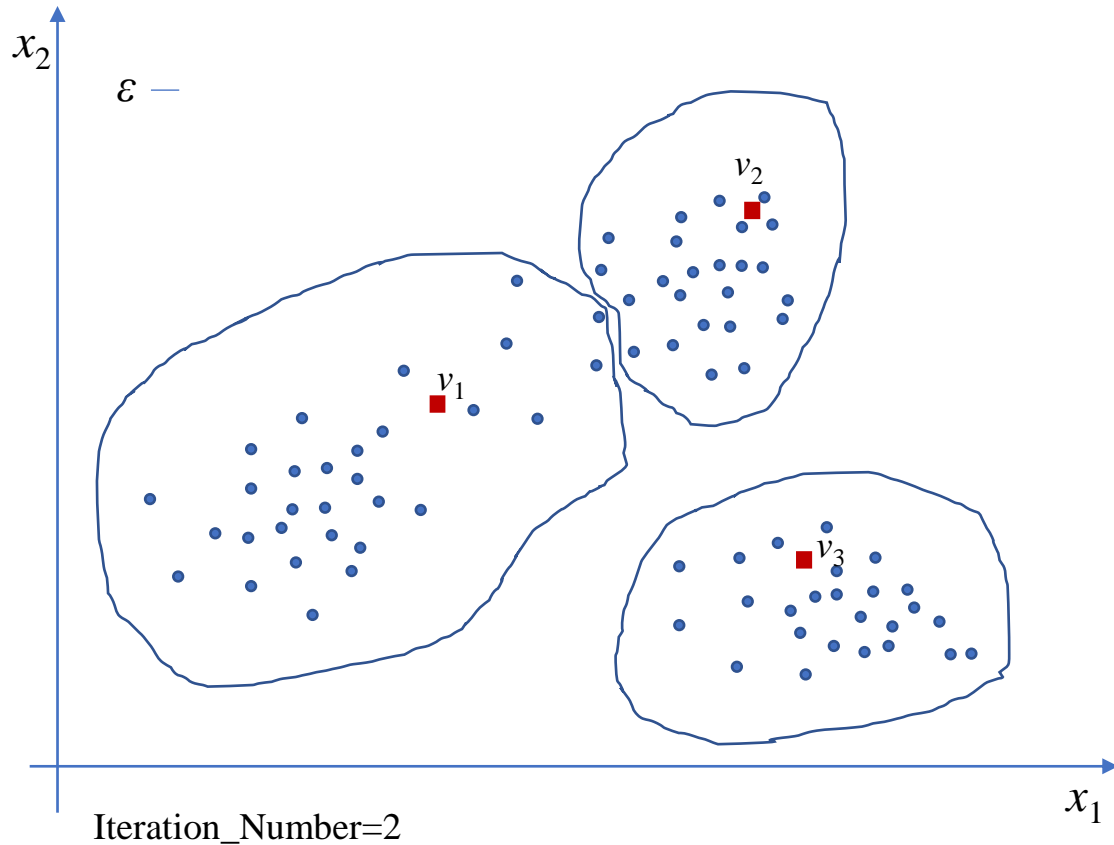
**Βήμα 2.** Δεδομένου ότι τα κέντρα των  $c$  συστάδων που είναι αποθηκευμένα στον πίνακα  $v$  είναι γνωστά υπολόγισε τον πίνακα των συναρτήσεων συμμετοχής  $u$  που έχουν τα δεδομένα στις  $c$  συστάδες

**Βήμα 3.** Δεδομένου ότι οι συναρτήσεις συμμετοχής των δεδομένων στις συστάδες, που είναι αποθηκευμένες στον πίνακα  $u$  είναι γνωστές, υπολόγισε τις καινούργιες τιμές των κέντρων  $v$  (δηλαδή οι τιμές του πίνακα  $v$  έχουν αλλάξει)

**Βήμα 4.** Για κάθε  $i=1, 2, \dots, c$  υπολογίζουμε την απόσταση  $d(i)=\sqrt{(v(i,1)-v\_old(i,1))^2+(v(i,2)-v\_old(i,2))^2+\dots+(v(i,p)-v\_old(i,p))^2}$  και υπολογίζουμε την μέγιστη από όλες αυτές:  $E=\max\{d(1), d(2), \dots, d(c)\}$

**EndWhile**

# Αλγόριθμος c-Means



Iteration\_Number=2

Βήμα 2. Δίνει τις συστάδες του σχήματος

Αρχικοποίηση των παραμέτρων

- Εισάγουμε τον πίνακα των δεδομένων  $x$  και προσδιορίζουμε την διάστασή του  $N \times p$
- Δίνουμε τιμή στον αριθμό των συστάδων  $c$  και μία πολύ μικρή τιμή στην παράμετρο  $\epsilon$
- Δίνουμε τυχαίες τιμές στα κέντρα των συστάδων  $v_1, v_2, \dots, v_c$  που αποθηκεύονται στον πίνακα  $v$ , όπου  $v_1=v(1,:)$ ,  $v_2=v(2,:)$ ,  $\dots$ ,  $v_c=v(c,:)$ . Άρα ο πίνακας  $v$  είναι  $c \times p$
- Δίνουμε μία τιμή στην παράμετρο  $E$  μεγαλύτερη από την τιμή του  $\epsilon$  (δηλαδή  $E > \epsilon$ )

Iteration\_Number=0

**While**  $E > \epsilon$  **do**

Iteration\_Number= Iteration\_Number+1

**Βήμα 1.** Αποθήκευσε τις τωρινές τιμές των κέντρων των συστάδων  $v$  σε έναν πίνακα  $v\_old$  που είναι βοηθητικός για να μην χαθούν αυτές οι τιμές των κέντρων:  $v\_old=v$

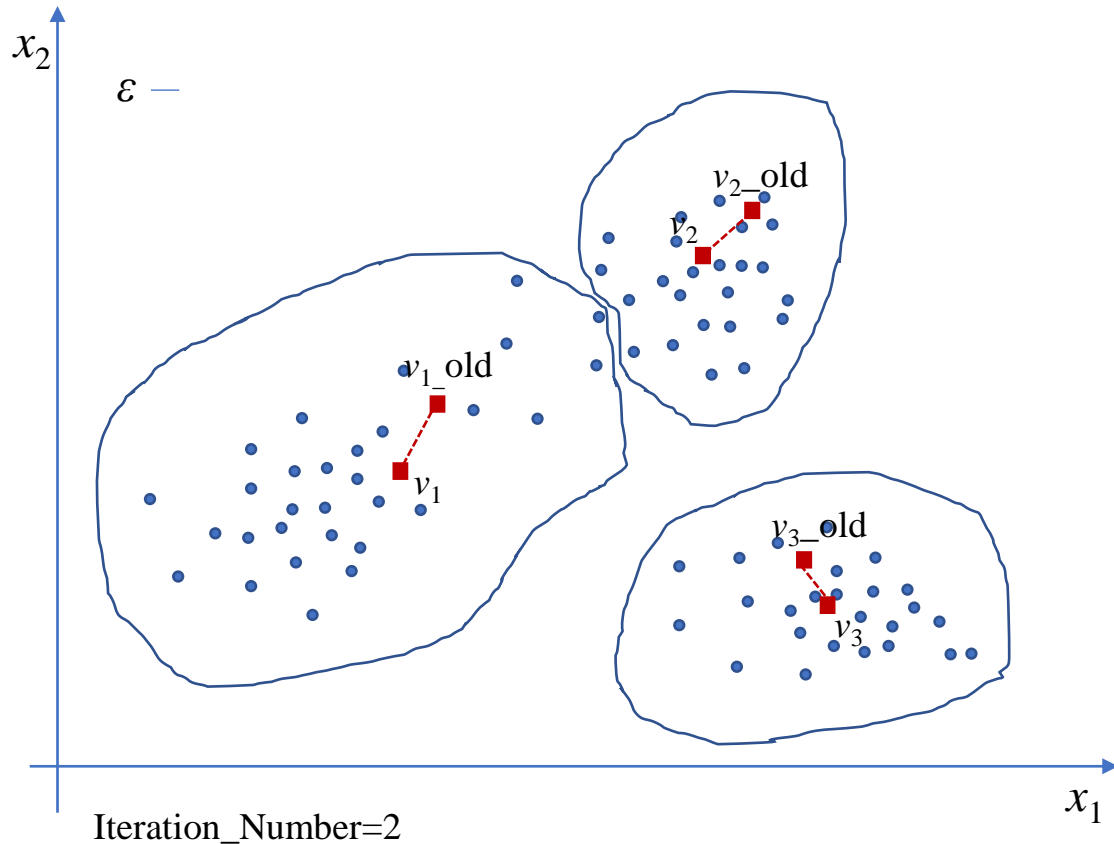
**Βήμα 2.** Δεδομένου ότι τα κέντρα των  $c$  συστάδων που είναι αποθηκευμένα στον πίνακα  $v$  είναι γνωστά υπολόγισε τον πίνακα των συναρτήσεων συμμετοχής  $u$  που έχουν τα δεδομένα στις  $c$  συστάδες

**Βήμα 3.** Δεδομένου ότι οι συναρτήσεις συμμετοχής των δεδομένων στις συστάδες, που είναι αποθηκευμένες στον πίνακα  $u$  είναι γνωστές, υπολόγισε τις καινούργιες τιμές των κέντρων  $v$  (δηλαδή οι τιμές του πίνακα  $v$  έχουν αλλάξει)

**Βήμα 4.** Για κάθε  $i=1, 2, \dots, c$  υπολογίζουμε την απόσταση  $d(i)=\sqrt{(v(i,1)-v\_old(i,1))^2+(v(i,2)-v\_old(i,2))^2+\dots+(v(i,p)-v\_old(i,p))^2}$  και υπολογίζουμε την μέγιστη από όλες αυτές:  $E=\max\{d(1), d(2), \dots, d(c)\}$

**EndWhile**

# Αλγόριθμος c-Means



Iteration\_Number=2

Βήμα 2. Δίνει τις συστάδες του σχήματος

Βήμα 3. Τα κέντρα μετακινούνται σε καινούργιες τιμές με κατεύθυνση προς τον κύριο όγκο δεδομένων σε κάθε συστάδα

Αρχικοποίηση των παραμέτρων

- Εισάγουμε τον πίνακα των δεδομένων  $x$  και προσδιορίζουμε την διάστασή του  $N \times p$
- Δίνουμε τιμή στον αριθμό των συστάδων  $c$  και μία πολύ μικρή τιμή στην παράμετρο  $\epsilon$
- Δίνουμε τυχαίες τιμές στα κέντρα των συστάδων  $v_1, v_2, \dots, v_c$  που αποθηκεύονται στον πίνακα  $v$ , όπου  $v_1=v(1,:)$ ,  $v_2=v(2,:)$ ,  $\dots$ ,  $v_c=v(c,:)$ . Άρα ο πίνακας  $v$  είναι  $c \times p$
- Δίνουμε μία τιμή στην παράμετρο  $E$  μεγαλύτερη από την τιμή του  $\epsilon$  (δηλαδή  $E > \epsilon$ )

Iteration\_Number=0

**While**  $E > \epsilon$  **do**

Iteration\_Number=Iteration\_Number+1

**Βήμα 1.** Αποθήκευσε τις τωρινές τιμές των κέντρων των συστάδων  $v$  σε έναν πίνακα  $v\_old$  που είναι βοηθητικός για να μην χαθούν αυτές οι τιμές των κέντρων:  $v\_old=v$

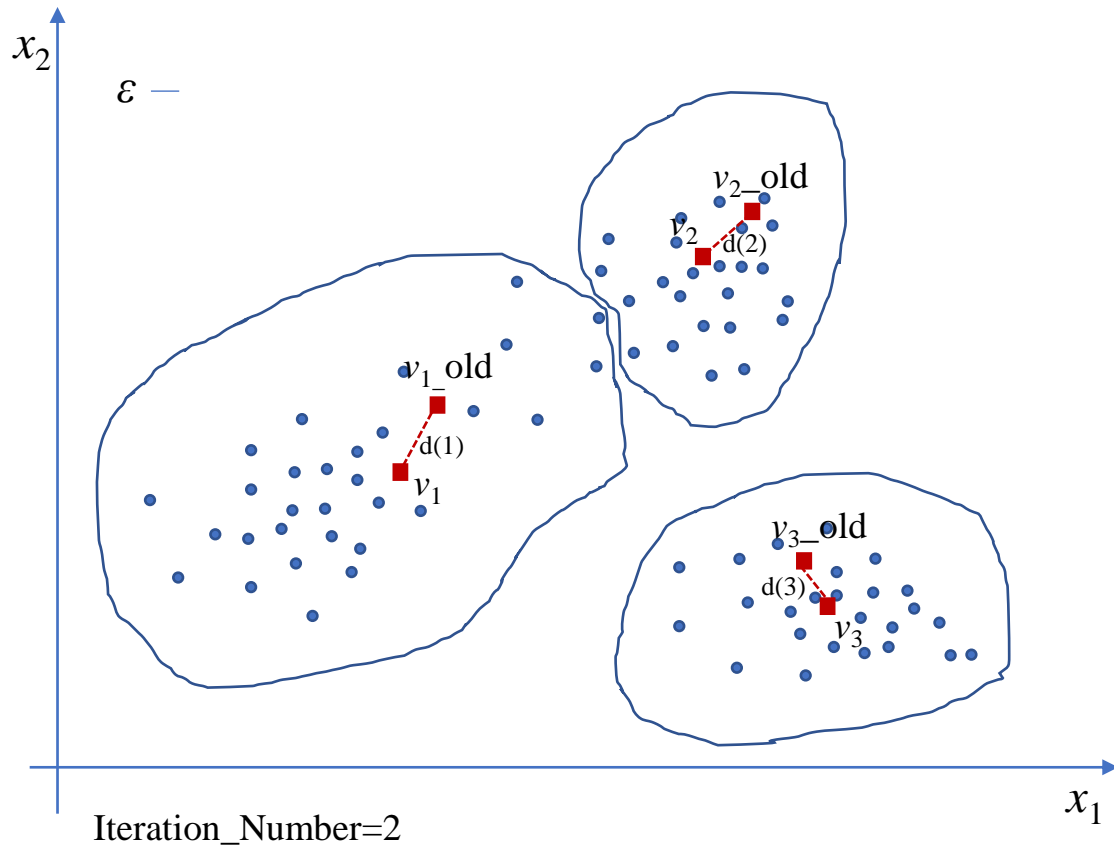
**Βήμα 2.** Δεδομένου ότι τα κέντρα των  $c$  συστάδων που είναι αποθηκευμένα στον πίνακα  $v$  είναι γνωστά υπολόγισε τον πίνακα των συναρτήσεων συμμετοχής  $u$  που έχουν τα δεδομένα στις  $c$  συστάδες

**Βήμα 3.** Δεδομένου ότι οι συναρτήσεις συμμετοχής των δεδομένων στις συστάδες, που είναι αποθηκευμένες στον πίνακα  $u$  είναι γνωστές, υπολόγισε τις καινούργιες τιμές των κέντρων  $v$  (δηλαδή οι τιμές του πίνακα  $v$  έχουν αλλάξει)

**Βήμα 4.** Για κάθε  $i=1, 2, \dots, c$  υπολογίζουμε την απόσταση  $d(i)=\sqrt{(v(i,1)-v\_old(i,1))^2+(v(i,2)-v\_old(i,2))^2+\dots+(v(i,p)-v\_old(i,p))^2}$  και υπολογίζουμε την μέγιστη από όλες αυτές:  
 $E=\max\{d(1), d(2), \dots, d(c)\}$

**EndWhile**

# Αλγόριθμος c-Means



Iteration\_Number=2

**Βήμα 2.** Δίνει τις συστάδες του σχήματος

**Βήμα 3.** Τα κέντρα μετακινούνται σε καινούργιες τιμές με κατεύθυνση προς τον κύριο όγκο δεδομένων σε κάθε συστάδα

**Βήμα 4.** Στο σχήμα φαίνονται οι αποστάσεις  $d(1)$ ,  $d(2)$  και  $d(3)$  και είναι προφανές ότι  $E=d(1)$ , το οποίο είναι  $E>\epsilon$

Αρχικοποίηση των παραμέτρων

- Εισάγουμε τον πίνακα των δεδομένων  $x$  και προσδιορίζουμε την διάστασή του  $N \times p$
- Δίνουμε τιμή στον αριθμό των συστάδων  $c$  και μία πολύ μικρή τιμή στην παράμετρο  $\epsilon$
- Δίνουμε τυχαίες τιμές στα κέντρα των συστάδων  $v_1, v_2, \dots, v_c$  που αποθηκεύονται στον πίνακα  $v$ , όπου  $v_1=v(1,:)$ ,  $v_2=v(2,:)$ ,  $\dots$ ,  $v_c=v(c,:)$ . Άρα ο πίνακας  $v$  είναι  $c \times p$
- Δίνουμε μία τιμή στην παράμετρο  $E$  μεγαλύτερη από την τιμή του  $\epsilon$  (δηλαδή  $E>\epsilon$ )

Iteration\_Number=0

**While**  $E>\epsilon$  **do**

Iteration\_Number=Iteration\_Number+1

**Βήμα 1.** Αποθήκευσε τις τωρινές τιμές των κέντρων των συστάδων  $v$  σε έναν πίνακα  $v\_old$  που είναι βοηθητικός για να μην χαθούν αυτές οι τιμές των κέντρων:  $v\_old=v$

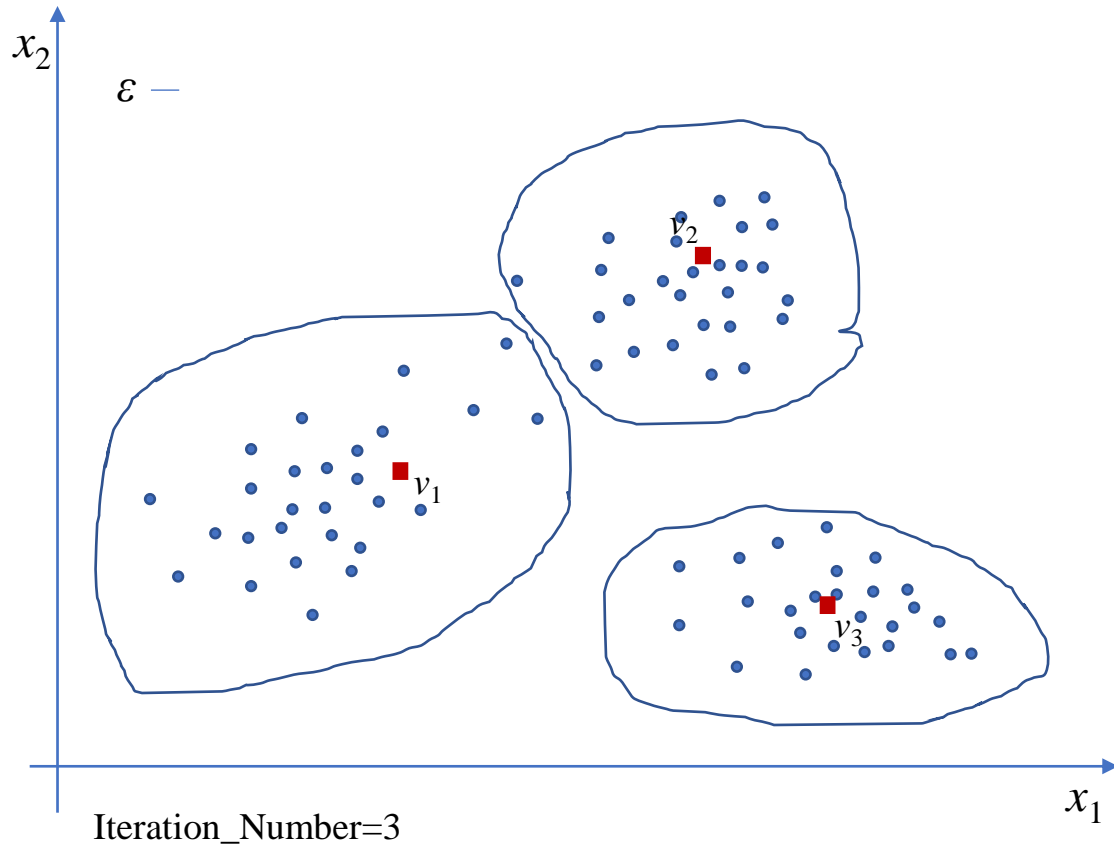
**Βήμα 2.** Δεδομένου ότι τα κέντρα των  $c$  συστάδων που είναι αποθηκευμένα στον πίνακα  $v$  είναι γνωστά υπολόγισε τον πίνακα των συναρτήσεων συμμετοχής  $u$  που έχουν τα δεδομένα στις  $c$  συστάδες

**Βήμα 3.** Δεδομένου ότι οι συναρτήσεις συμμετοχής των δεδομένων στις συστάδες, που είναι αποθηκευμένες στον πίνακα  $u$  είναι γνωστές, υπολόγισε τις καινούργιες τιμές των κέντρων  $v$  (δηλαδή οι τιμές του πίνακα  $v$  έχουν αλλάξει)

**Βήμα 4.** Για κάθε  $i=1, 2, \dots, c$  υπολογίζουμε την απόσταση  $d(i)=\sqrt{(v(i,1)-v\_old(i,1))^2+(v(i,2)-v\_old(i,2))^2+\dots+(v(i,p)-v\_old(i,p))^2}$  και υπολογίζουμε την μέγιστη από όλες αυτές:  $E=\max\{d(1), d(2), \dots, d(c)\}$

**EndWhile**

# Αλγόριθμος c-Means



Iteration\_Number=3

Βήμα 2. Δίνει τις συστάδες του σχήματος

Αρχικοποίηση των παραμέτρων

- Εισάγουμε τον πίνακα των δεδομένων  $x$  και προσδιορίζουμε την διάστασή του  $N \times p$
- Δίνουμε τιμή στον αριθμό των συστάδων  $c$  και μία πολύ μικρή τιμή στην παράμετρο  $\epsilon$
- Δίνουμε τυχαίες τιμές στα κέντρα των συστάδων  $v_1, v_2, \dots, v_c$  που αποθηκεύονται στον πίνακα  $v$ , όπου  $v_1=v(1,:)$ ,  $v_2=v(2,:)$ ,  $\dots$ ,  $v_c=v(c,:)$ . Άρα ο πίνακας  $v$  είναι  $c \times p$
- Δίνουμε μία τιμή στην παράμετρο  $E$  μεγαλύτερη από την τιμή του  $\epsilon$  (δηλαδή  $E > \epsilon$ )

Iteration\_Number=0

**While**  $E > \epsilon$  **do**

Iteration\_Number=Iteration\_Number+1

**Βήμα 1.** Αποθήκευσε τις τωρινές τιμές των κέντρων των συστάδων  $v$  σε έναν πίνακα  $v\_old$  που είναι βοηθητικός για να μην χαθούν αυτές οι τιμές των κέντρων:  $v\_old=v$

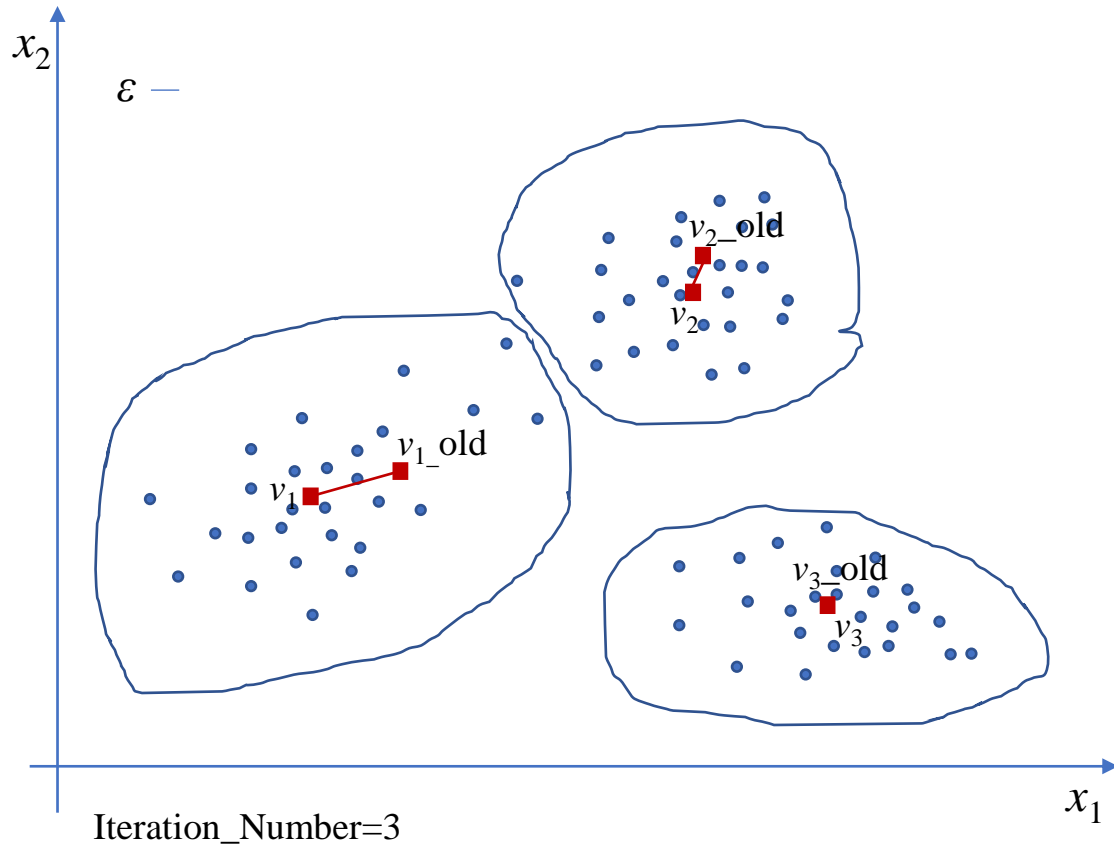
**Βήμα 2.** Δεδομένου ότι τα κέντρα των  $c$  συστάδων που είναι αποθηκευμένα στον πίνακα  $v$  είναι γνωστά υπολόγισε τον πίνακα των συναρτήσεων συμμετοχής  $u$  που έχουν τα δεδομένα στις  $c$  συστάδες

**Βήμα 3.** Δεδομένου ότι οι συναρτήσεις συμμετοχής των δεδομένων στις συστάδες, που είναι αποθηκευμένες στον πίνακα  $u$  είναι γνωστές, υπολόγισε τις καινούργιες τιμές των κέντρων  $v$  (δηλαδή οι τιμές του πίνακα  $v$  έχουν αλλάξει)

**Βήμα 4.** Για κάθε  $i=1, 2, \dots, c$  υπολογίζουμε την απόσταση  $d(i)=\sqrt{(v(i,1)-v\_old(i,1))^2+(v(i,2)-v\_old(i,2))^2+\dots+(v(i,p)-v\_old(i,p))^2}$  και υπολογίζουμε την μέγιστη από όλες αυτές:  $E=\max\{d(1), d(2), \dots, d(c)\}$

**EndWhile**

# Αλγόριθμος c-Means



Iteration\_Number=3

Βήμα 2. Δίνει τις συστάδες του σχήματος

Βήμα 3. Τα κέντρα μετακινούνται σε καινούργιες τιμές με κατεύθυνση προς τον κύριο όγκο δεδομένων σε κάθε συστάδα

Αρχικοποίηση των παραμέτρων

- Εισάγουμε τον πίνακα των δεδομένων  $x$  και προσδιορίζουμε την διάστασή του  $N \times p$
- Δίνουμε τιμή στον αριθμό των συστάδων  $c$  και μία πολύ μικρή τιμή στην παράμετρο  $\epsilon$
- Δίνουμε τυχαίες τιμές στα κέντρα των συστάδων  $v_1, v_2, \dots, v_c$  που αποθηκεύονται στον πίνακα  $v$ , όπου  $v_1=v(1,:)$ ,  $v_2=v(2,:)$ ,  $\dots$ ,  $v_c=v(c,:)$ . Άρα ο πίνακας  $v$  είναι  $c \times p$
- Δίνουμε μία τιμή στην παράμετρο  $E$  μεγαλύτερη από την τιμή του  $\epsilon$  (δηλαδή  $E > \epsilon$ )

Iteration\_Number=0

**While**  $E > \epsilon$  **do**

Iteration\_Number=Iteration\_Number+1

**Βήμα 1.** Αποθήκευσε τις τωρινές τιμές των κέντρων των συστάδων  $v$  σε έναν πίνακα  $v\_old$  που είναι βοηθητικός για να μην χαθούν αυτές οι τιμές των κέντρων:  $v\_old=v$

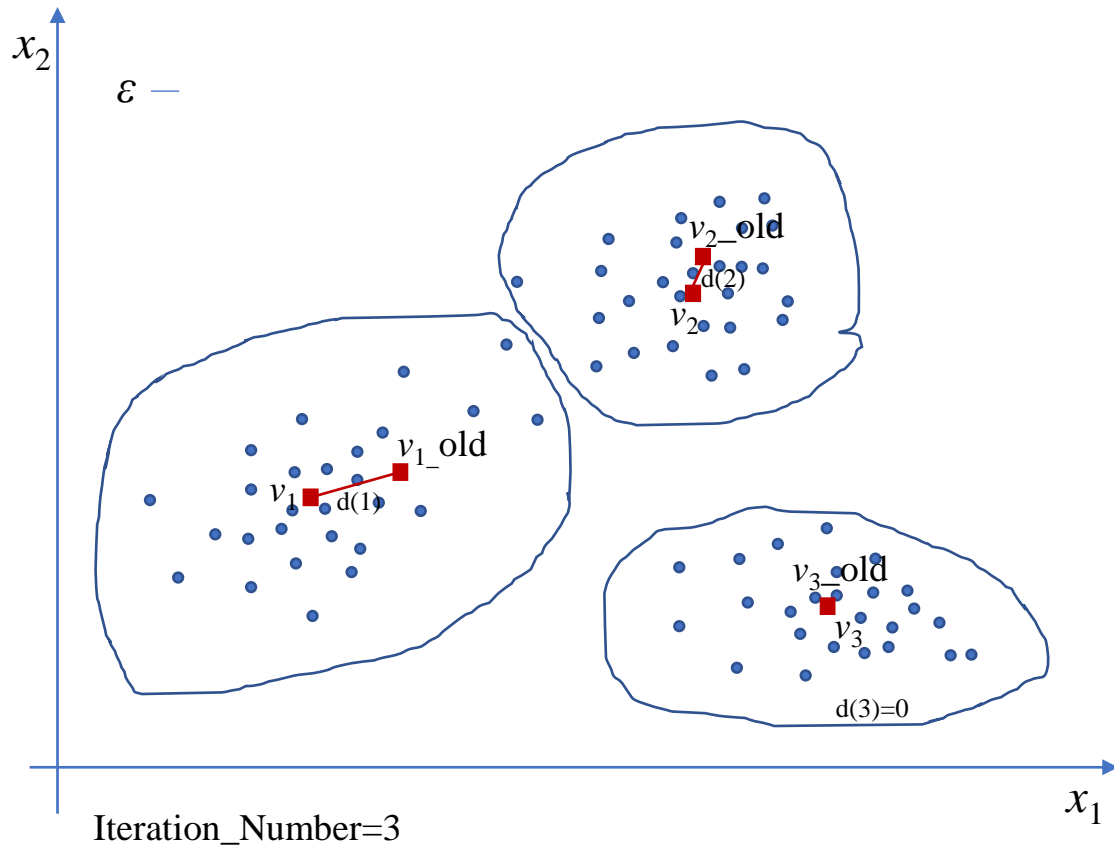
**Βήμα 2.** Δεδομένου ότι τα κέντρα των  $c$  συστάδων που είναι αποθηκευμένα στον πίνακα  $v$  είναι γνωστά υπολόγισε τον πίνακα των συναρτήσεων συμμετοχής  $u$  που έχουν τα δεδομένα στις  $c$  συστάδες

**Βήμα 3.** Δεδομένου ότι οι συναρτήσεις συμμετοχής των δεδομένων στις συστάδες, που είναι αποθηκευμένες στον πίνακα  $u$  είναι γνωστές, υπολόγισε τις καινούργιες τιμές των κέντρων  $v$  (δηλαδή οι τιμές του πίνακα  $v$  έχουν αλλάξει)

**Βήμα 4.** Για κάθε  $i=1, 2, \dots, c$  υπολογίζουμε την απόσταση  $d(i)=\sqrt{(v(i,1)-v\_old(i,1))^2+(v(i,2)-v\_old(i,2))^2+\dots+(v(i,p)-v\_old(i,p))^2}$  και υπολογίζουμε την μέγιστη από όλες αυτές:  $E=\max\{d(1), d(2), \dots, d(c)\}$

**EndWhile**

# Αλγόριθμος c-Means



Iteration\_Number=3

Βήμα 2. Δίνει τις συστάδες του σχήματος

Βήμα 3. Τα κέντρα μετακινούνται σε καινούργιες τιμές με κατεύθυνση προς τον κύριο όγκο δεδομένων σε κάθε συστάδα

Βήμα 4. Στο σχήμα φαίνονται οι αποστάσεις  $d(1)$ ,  $d(2)$  και  $d(3)=0$  και είναι προφανές ότι  $E=d(1)$ , το οποίο είναι  $E>\epsilon$

Αρχικοποίηση των παραμέτρων

- Εισάγουμε τον πίνακα των δεδομένων  $x$  και προσδιορίζουμε την διάστασή του  $N \times p$
- Δίνουμε τιμή στον αριθμό των συστάδων  $c$  και μία πολύ μικρή τιμή στην παράμετρο  $\epsilon$
- Δίνουμε τυχαίες τιμές στα κέντρα των συστάδων  $v_1, v_2, \dots, v_c$  που αποθηκεύονται στον πίνακα  $v$ , όπου  $v_1=v(1,:)$ ,  $v_2=v(2,:)$ ,  $\dots$ ,  $v_c=v(c,:)$ . Άρα ο πίνακας  $v$  είναι  $c \times p$
- Δίνουμε μία τιμή στην παράμετρο  $E$  μεγαλύτερη από την τιμή του  $\epsilon$  (δηλαδή  $E>\epsilon$ )

Iteration\_Number=0

**While**  $E>\epsilon$  **do**

Iteration\_Number=Iteration\_Number+1

**Βήμα 1.** Αποθήκευσε τις τωρινές τιμές των κέντρων των συστάδων  $v$  σε έναν πίνακα  $v\_old$  που είναι βοηθητικός για να μην χαθούν αυτές οι τιμές των κέντρων:  $v\_old=v$

**Βήμα 2.** Δεδομένου ότι τα κέντρα των  $c$  συστάδων που είναι αποθηκευμένα στον πίνακα  $v$  είναι γνωστά υπολόγισε τον πίνακα των συναρτήσεων συμμετοχής  $u$  που έχουν τα δεδομένα στις  $c$  συστάδες

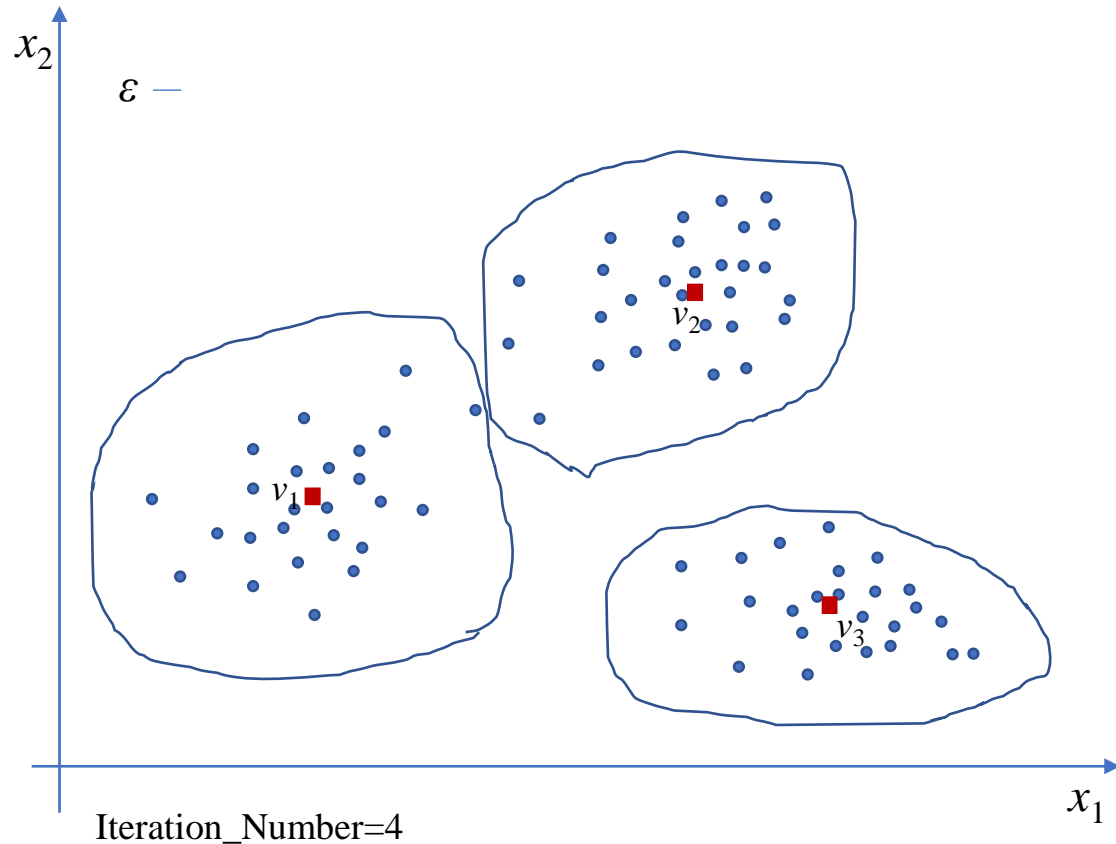
**Βήμα 3.** Δεδομένου ότι οι συναρτήσεις συμμετοχής των δεδομένων στις συστάδες, που είναι αποθηκευμένες στον πίνακα  $u$  είναι γνωστές, υπολόγισε τις καινούργιες τιμές των κέντρων  $v$  (δηλαδή οι τιμές του πίνακα  $v$  έχουν αλλάξει)

**Βήμα 4.** Για κάθε  $i=1, 2, \dots, c$  υπολογίζουμε την απόσταση  $d(i)=\sqrt{(v(i,1)-v\_old(i,1))^2+(v(i,2)-v\_old(i,2))^2+\dots+(v(i,p)-v\_old(i,p))^2}$  και υπολογίζουμε την μέγιστη από όλες αυτές:  $E=\max\{d(1), d(2), \dots, d(c)\}$

**EndWhile**



# Αλγόριθμος c-Means



Iteration\_Number=4

Βήμα 2. Δίνει τις συστάδες του σχήματος

Αρχικοποίηση των παραμέτρων

- Εισάγουμε τον πίνακα των δεδομένων  $x$  και προσδιορίζουμε την διάστασή του  $N \times p$
- Δίνουμε τιμή στον αριθμό των συστάδων  $c$  και μία πολύ μικρή τιμή στην παράμετρο  $\epsilon$
- Δίνουμε τυχαίες τιμές στα κέντρα των συστάδων  $v_1, v_2, \dots, v_c$  που αποθηκεύονται στον πίνακα  $v$ , όπου  $v_1=v(1,:)$ ,  $v_2=v(2,:)$ ,  $\dots$ ,  $v_c=v(c,:)$ . Άρα ο πίνακας  $v$  είναι  $c \times p$
- Δίνουμε μία τιμή στην παράμετρο  $E$  μεγαλύτερη από την τιμή του  $\epsilon$  (δηλαδή  $E > \epsilon$ )

Iteration\_Number=0

**While**  $E > \epsilon$  **do**

Iteration\_Number=Iteration\_Number+1

**Βήμα 1.** Αποθήκευσε τις τωρινές τιμές των κέντρων των συστάδων  $v$  σε έναν πίνακα  $v\_old$  που είναι βοηθητικός για να μην χαθούν αυτές οι τιμές των κέντρων:  $v\_old=v$

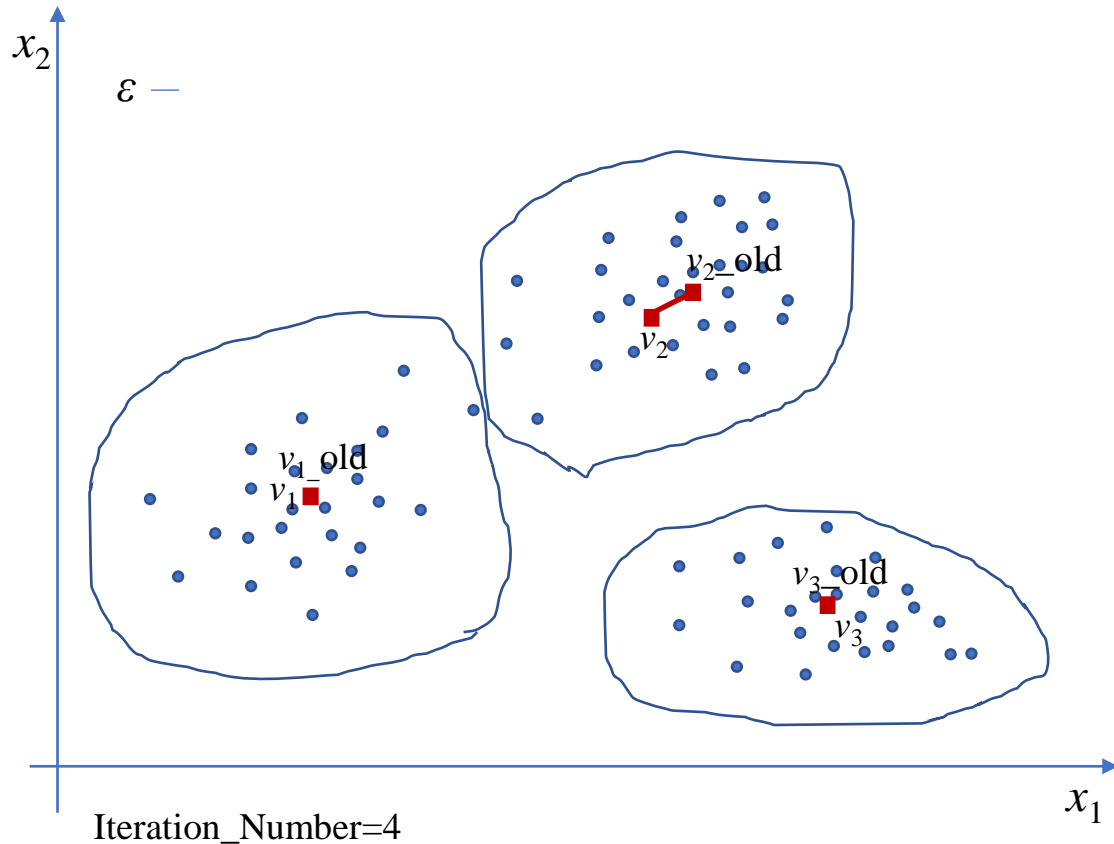
**Βήμα 2.** Δεδομένου ότι τα κέντρα των  $c$  συστάδων που είναι αποθηκευμένα στον πίνακα  $v$  είναι γνωστά υπολόγισε τον πίνακα των συναρτήσεων συμμετοχής  $u$  που έχουν τα δεδομένα στις  $c$  συστάδες

**Βήμα 3.** Δεδομένου ότι οι συναρτήσεις συμμετοχής των δεδομένων στις συστάδες, που είναι αποθηκευμένες στον πίνακα  $u$  είναι γνωστές, υπολόγισε τις καινούργιες τιμές των κέντρων  $v$  (δηλαδή οι τιμές του πίνακα  $v$  έχουν αλλάξει)

**Βήμα 4.** Για κάθε  $i=1, 2, \dots, c$  υπολογίζουμε την απόσταση  $d(i)=\sqrt{(v(i,1)-v\_old(i,1))^2+(v(i,2)-v\_old(i,2))^2+\dots+(v(i,p)-v\_old(i,p))^2}$  και υπολογίζουμε την μέγιστη από όλες αυτές:  $E=\max\{d(1), d(2), \dots, d(c)\}$

**EndWhile**

# Αλγόριθμος c-Means



Iteration\_Number=4

Βήμα 2. Δίνει τις συστάδες του σχήματος

Βήμα 3. Τα κέντρα μετακινούνται σε καινούργιες τιμές με κατεύθυνση προς τον κύριο όγκο δεδομένων σε κάθε συστάδα

Αρχικοποίηση των παραμέτρων

- Εισάγουμε τον πίνακα των δεδομένων  $x$  και προσδιορίζουμε την διάστασή του  $N \times p$
- Δίνουμε τιμή στον αριθμό των συστάδων  $c$  και μία πολύ μικρή τιμή στην παράμετρο  $\epsilon$
- Δίνουμε τυχαίες τιμές στα κέντρα των συστάδων  $v_1, v_2, \dots, v_c$  που αποθηκεύονται στον πίνακα  $v$ , όπου  $v_1=v(1,:)$ ,  $v_2=v(2,:)$ ,  $\dots$ ,  $v_c=v(c,:)$ . Άρα ο πίνακας  $v$  είναι  $c \times p$
- Δίνουμε μία τιμή στην παράμετρο  $E$  μεγαλύτερη από την τιμή του  $\epsilon$  (δηλαδή  $E > \epsilon$ )

Iteration\_Number=0

**While**  $E > \epsilon$  **do**

Iteration\_Number=Iteration\_Number+1

**Βήμα 1.** Αποθήκευσε τις τωρινές τιμές των κέντρων των συστάδων  $v$  σε έναν πίνακα  $v\_old$  που είναι βοηθητικός για να μην χαθούν αυτές οι τιμές των κέντρων:  $v\_old=v$

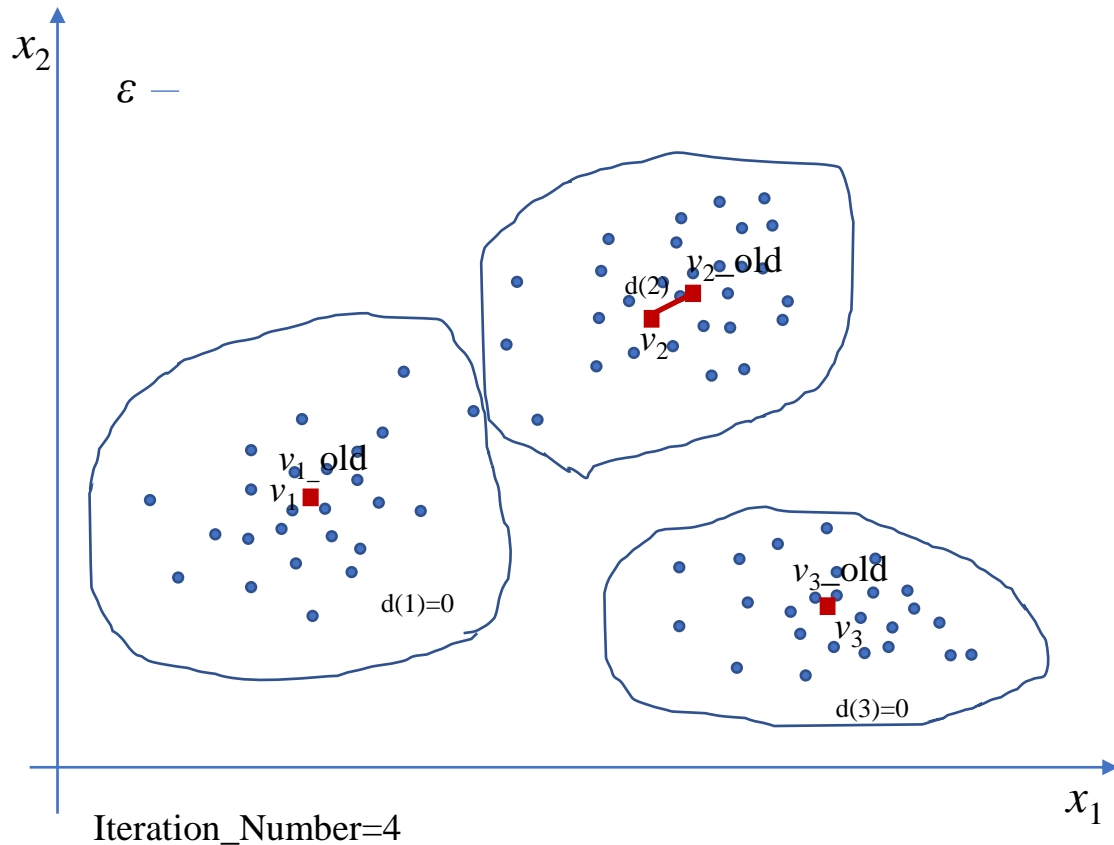
**Βήμα 2.** Δεδομένου ότι τα κέντρα των  $c$  συστάδων που είναι αποθηκευμένα στον πίνακα  $v$  είναι γνωστά υπολόγισε τον πίνακα των συναρτήσεων συμμετοχής  $u$  που έχουν τα δεδομένα στις  $c$  συστάδες

**Βήμα 3.** Δεδομένου ότι οι συναρτήσεις συμμετοχής των δεδομένων στις συστάδες, που είναι αποθηκευμένες στον πίνακα  $u$  είναι γνωστές, υπολόγισε τις καινούργιες τιμές των κέντρων  $v$  (δηλαδή οι τιμές του πίνακα  $v$  έχουν αλλάξει)

**Βήμα 4.** Για κάθε  $i=1, 2, \dots, c$  υπολογίζουμε την απόσταση  $d(i)=\sqrt{(v(i,1)-v\_old(i,1))^2+(v(i,2)-v\_old(i,2))^2+\dots+(v(i,p)-v\_old(i,p))^2}$  και υπολογίζουμε την μέγιστη από όλες αυτές:  $E=\max\{d(1), d(2), \dots, d(c)\}$

**EndWhile**

# Αλγόριθμος c-Means



Iteration\_Number=4

Βήμα 2. Δίνει τις συστάδες του σχήματος

Βήμα 3. Τα κέντρα μετακινούνται σε καινούργιες τιμές με κατεύθυνση προς τον κύριο όγκο δεδομένων σε κάθε συστάδα

Βήμα 4. Στο σχήμα φαίνονται οι αποστάσεις  $d(1)=0$ ,  $d(2)$  και  $d(3)=0$  και είναι προφανές ότι  $E=d(2)$ , το οποίο είναι  $E>\epsilon$

Αρχικοποίηση των παραμέτρων

- Εισάγουμε τον πίνακα των δεδομένων  $x$  και προσδιορίζουμε την διάστασή του  $N \times p$
- Δίνουμε τιμή στον αριθμό των συστάδων  $c$  και μία πολύ μικρή τιμή στην παράμετρο  $\epsilon$
- Δίνουμε τυχαίες τιμές στα κέντρα των συστάδων  $v_1, v_2, \dots, v_c$  που αποθηκεύονται στον πίνακα  $v$ , όπου  $v_1=v(1,:)$ ,  $v_2=v(2,:)$ ,  $\dots$ ,  $v_c=v(c,:)$ . Άρα ο πίνακας  $v$  είναι  $c \times p$
- Δίνουμε μία τιμή στην παράμετρο  $E$  μεγαλύτερη από την τιμή του  $\epsilon$  (δηλαδή  $E>\epsilon$ )

Iteration\_Number=0

**While**  $E>\epsilon$  **do**

Iteration\_Number=Iteration\_Number+1

**Βήμα 1.** Αποθήκευσε τις τωρινές τιμές των κέντρων των συστάδων  $v$  σε έναν πίνακα  $v\_old$  που είναι βοηθητικός για να μην χαθούν αυτές οι τιμές των κέντρων:  $v\_old=v$

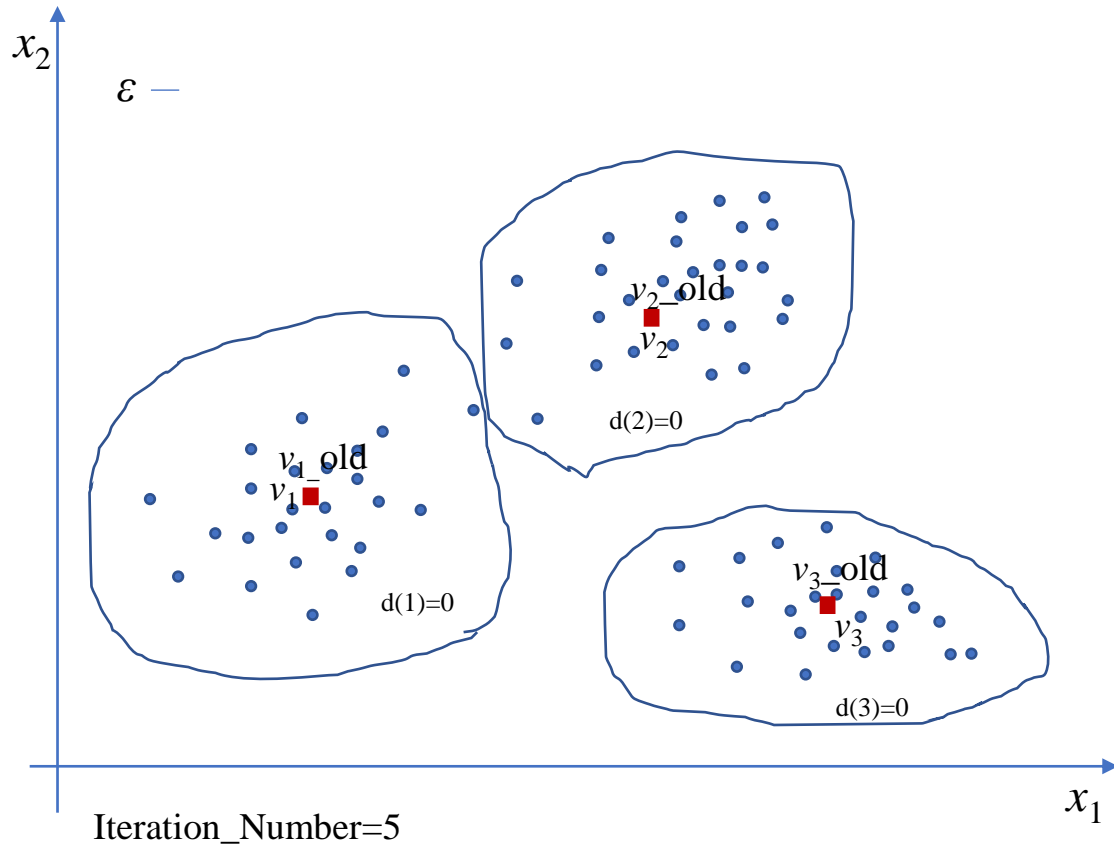
**Βήμα 2.** Δεδομένου ότι τα κέντρα των  $c$  συστάδων που είναι αποθηκευμένα στον πίνακα  $v$  είναι γνωστά υπολόγισε τον πίνακα των συναρτήσεων συμμετοχής  $u$  που έχουν τα δεδομένα στις  $c$  συστάδες

**Βήμα 3.** Δεδομένου ότι οι συναρτήσεις συμμετοχής των δεδομένων στις συστάδες, που είναι αποθηκευμένες στον πίνακα  $u$  είναι γνωστές, υπολόγισε τις καινούργιες τιμές των κέντρων  $v$  (δηλαδή οι τιμές του πίνακα  $v$  έχουν αλλάξει)

**Βήμα 4.** Για κάθε  $i=1, 2, \dots, c$  υπολογίζουμε την απόσταση  $d(i)=\sqrt{(v(i,1)-v\_old(i,1))^2+(v(i,2)-v\_old(i,2))^2+\dots+(v(i,p)-v\_old(i,p))^2}$  και υπολογίζουμε την μέγιστη από όλες αυτές:  $E=\max\{d(1), d(2), \dots, d(c)\}$

**EndWhile**

# Αλγόριθμος c-Means



Iteration\_Number=5

Βήμα 2. Δίνει τις συστάδες του σχήματος

Βήμα 3. Τα κέντρα μετακινούνται σε καινούργιες τιμές με κατεύθυνση προς τον κύριο όγκο δεδομένων σε κάθε συστάδα

Βήμα 4. Στο σχήμα φαίνονται οι αποστάσεις  $d(1)=0$ ,  $d(2)=0$  και  $d(3)=0$  και είναι προφανές ότι  $E=0$ , το οποίο είναι  $E < \epsilon$

Αρχικοποίηση των παραμέτρων

- Εισάγουμε τον πίνακα των δεδομένων  $x$  και προσδιορίζουμε την διάστασή του  $N \times p$
- Δίνουμε τιμή στον αριθμό των συστάδων  $c$  και μία πολύ μικρή τιμή στην παράμετρο  $\epsilon$
- Δίνουμε τυχαίες τιμές στα κέντρα των συστάδων  $v_1, v_2, \dots, v_c$  που αποθηκεύονται στον πίνακα  $v$ , όπου  $v_1=v(1,:)$ ,  $v_2=v(2,:)$ ,  $\dots$ ,  $v_c=v(c,:)$ . Άρα ο πίνακας  $v$  είναι  $c \times p$
- Δίνουμε μία τιμή στην παράμετρο  $E$  μεγαλύτερη από την τιμή του  $\epsilon$  (δηλαδή  $E > \epsilon$ )

Iteration\_Number=0

**While**  $E > \epsilon$  **do**

Iteration\_Number=Iteration\_Number+1

**Βήμα 1.** Αποθήκευσε τις τωρινές τιμές των κέντρων των συστάδων  $v$  σε έναν πίνακα  $v\_old$  που είναι βοηθητικός για να μην χαθούν αυτές οι τιμές των κέντρων:  $v\_old=v$

**Βήμα 2.** Δεδομένου ότι τα κέντρα των  $c$  συστάδων που είναι αποθηκευμένα στον πίνακα  $v$  είναι γνωστά υπολόγισε τον πίνακα των συναρτήσεων συμμετοχής  $u$  που έχουν τα δεδομένα στις  $c$  συστάδες

**Βήμα 3.** Δεδομένου ότι οι συναρτήσεις συμμετοχής των δεδομένων στις συστάδες, που είναι αποθηκευμένες στον πίνακα  $u$  είναι γνωστές, υπολόγισε τις καινούργιες τιμές των κέντρων  $v$  (δηλαδή οι τιμές του πίνακα  $v$  έχουν αλλάξει)

**Βήμα 4.** Για κάθε  $i=1, 2, \dots, c$  υπολογίζουμε την απόσταση  $d(i)=\sqrt{(v(i,1)-v\_old(i,1))^2+(v(i,2)-v\_old(i,2))^2+\dots+(v(i,p)-v\_old(i,p))^2}$  και υπολογίζουμε την μέγιστη από όλες αυτές:  $E=\max\{d(1), d(2), \dots, d(c)\}$

**EndWhile**

ΚΑΛΟ ΑΠΟΓΕΥΜΑ