

Μάθημα 5

ΕΞΟΥΥΞΗ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΨΗΦΙΑΚΟ ΚΑΙ ΔΙΑΔΙΚΤΥΑΚΟ ΠΕΡΙΕΧΟΜΕΝΟ

Συσταδοποίηση Δεδομένων

Στατιστική Ανάλυση Δεδομένων

Μονοδιάστατα Δεδομένα

1. Μέσος Όρος

	x
x_1	5
x_2	4
x_3	-3
x_4	7
x_5	9
x_6	-10

Στατιστική Ανάλυση Δεδομένων

Μονοδιάστατα Δεδομένα

1. Μέσος Όρος

	x
x_1	5
x_2	4
x_3	-3
x_4	7
x_5	9
x_6	-10



Στατιστική Ανάλυση Δεδομένων

Μονοδιάστατα Δεδομένα

1. Μέσος Όρος

	x
x_1	5
x_2	4
x_3	-3
x_4	7
x_5	9
x_6	-10



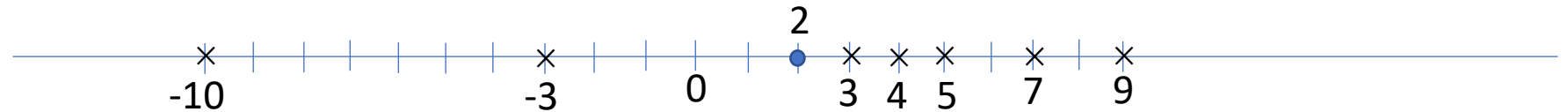
$$\mu = \frac{\sum_{k=1}^6 x_k}{6} = \frac{5 + 4 + (-3) + 7 + 9 + (-10)}{6} = 2$$

Στατιστική Ανάλυση Δεδομένων

Μονοδιάστατα Δεδομένα

1. Μέσος Όρος

	x
x_1	5
x_2	4
x_3	-3
x_4	7
x_5	9
x_6	-10



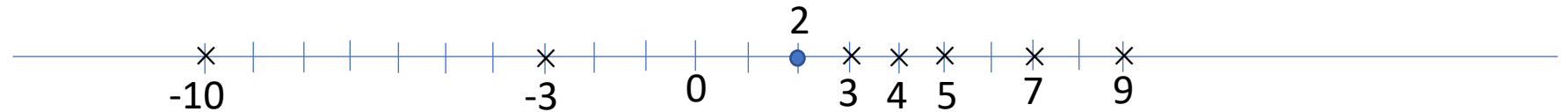
$$\mu = \frac{\sum_{k=1}^6 x_k}{6} = \frac{5 + 4 + (-3) + 7 + 9 + (-10)}{6} = 2$$

Στατιστική Ανάλυση Δεδομένων

Μονοδιάστατα Δεδομένα

1. Μέσος Όρος

	x
x_1	5
x_2	4
x_3	-3
x_4	7
x_5	9
x_6	-10



$$\mu = \frac{\sum_{k=1}^6 x_k}{6} = \frac{5 + 4 + (-3) + 7 + 9 + (-10)}{6} = 2$$

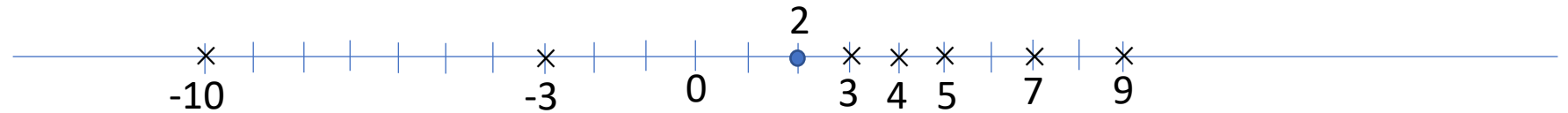
$$\mu = \frac{\sum_{k=1}^N x_k}{N}$$

Στατιστική Ανάλυση Δεδομένων

Μονοδιάστατα Δεδομένα

1. Μέσος Όρος

	x
x_1	5
x_2	4
x_3	-3
x_4	7
x_5	9
x_6	-10



$$\mu = \frac{\sum_{k=1}^6 x_k}{6} = \frac{5 + 4 + (-3) + 7 + 9 + (-10)}{6} = 2$$

$$\mu = \frac{\sum_{k=1}^N x_k}{N}$$

```
x=[5; 4; -3; 7; 9; -10];
```

```
N=6;
```

```
s=0;
```

```
for i=1:N
```

```
    s=s+x(i);
```

```
endfor
```

```
MO=s/N;
```

```
MO=mean(x);
```

Στατιστική Ανάλυση Δεδομένων

Μονοδιάστατα Δεδομένα

2. Διάμεσος (Ορίζεται μόνο για Μονοδιάστατα Δεδομένα)

Για έναν πίνακα x με N αριθμούς ο διάμεσος υπολογίζεται ως εξής:

Τα στοιχεία του πίνακα μπαίνουν σε αύξουσα ή φθίνουσα σειρά

Στατιστική Ανάλυση Δεδομένων

Μονοδιάστατα Δεδομένα

2. Διάμεσος (Ορίζεται μόνο για Μονοδιάστατα Δεδομένα)

Για έναν πίνακα x με N αριθμούς ο διάμεσος υπολογίζεται ως εξής:

Τα στοιχεία του πίνακα μπαίνουν σε αύξουσα ή φθίνουσα σειρά

Αν N περιττός τότε $\Delta(x) = x_{(N+1)/2}$

Αν N άρτιος τότε $\Delta(x) = \frac{x_{N/2} + x_{(N/2)+1}}{2}$

$D = \text{median}(x)$;

Στατιστική Ανάλυση Δεδομένων

Μονοδιάστατα Δεδομένα

2. Διάμεσος (Ορίζεται μόνο για Μονοδιάστατα Δεδομένα)

Για έναν πίνακα x με N αριθμούς ο διάμεσος υπολογίζεται ως εξής:

Τα στοιχεία του πίνακα μπαίνουν σε αύξουσα ή φθίνουσα σειρά

Αν N περιττός τότε $\Delta(x) = x_{(N+1)/2}$

Αν N άρτιος τότε $\Delta(x) = \frac{x_{N/2} + x_{(N/2)+1}}{2}$

} $D = \text{median}(x)$;

5
4
-3
7
9
-10

Στατιστική Ανάλυση Δεδομένων

Μονοδιάστατα Δεδομένα

2. Διάμεσος (Ορίζεται μόνο για Μονοδιάστατα Δεδομένα)

Για έναν πίνακα x με N αριθμούς ο διάμεσος υπολογίζεται ως εξής:


Τα στοιχεία του πίνακα μπαίνουν σε αύξουσα ή φθίνουσα σειρά

Αν N περιττός τότε $\Delta(x) = x_{(N+1)/2}$

Αν N άρτιος τότε $\Delta(x) = \frac{x_{N/2} + x_{(N/2)+1}}{2}$

} $D = \text{median}(x)$;

5	9
4	7
-3	5
7	4
9	-3
-10	-10



Στατιστική Ανάλυση Δεδομένων

Μονοδιάστατα Δεδομένα

2. Διάμεσος (Ορίζεται μόνο για Μονοδιάστατα Δεδομένα)

Για έναν πίνακα x με N αριθμούς ο διάμεσος υπολογίζεται ως εξής:

Τα στοιχεία του πίνακα μπαίνουν σε αύξουσα ή φθίνουσα σειρά

Αν N περιττός τότε $\Delta(x) = x_{(N+1)/2}$

Αν N άρτιος τότε $\Delta(x) = \frac{x_{N/2} + x_{(N/2)+1}}{2}$

$D = \text{median}(x)$;

5
4
-3
7
9
-10



9
7
5
4
-3
-10



$$\Delta(x) = (5 + 4) / 2 = 4.5$$

Στατιστική Ανάλυση Δεδομένων

Μονοδιάστατα Δεδομένα

2. Διάμεσος (Ορίζεται μόνο για Μονοδιάστατα Δεδομένα)

Για έναν πίνακα x με N αριθμούς ο διάμεσος υπολογίζεται ως εξής:

Τα στοιχεία του πίνακα μπαίνουν σε αύξουσα ή φθίνουσα σειρά

Αν N περιττός τότε $\Delta(x) = x_{(N+1)/2}$

Αν N άρτιος τότε $\Delta(x) = \frac{x_{N/2} + x_{(N/2)+1}}{2}$

$D = \text{median}(x)$;

5
4
-3
7
9
-10



9
7
5
4
-3
-10



$$\Delta(x) = (5 + 4) / 2 = 4.5$$

9
7
5
4
-3
-10
-11



$$\Delta(x) = 4$$

Στατιστική Ανάλυση Δεδομένων

Μονοδιάστατα Δεδομένα

3. Απόκλιση και Τυπική Απόκλιση

Μέσος Όρος

$$\mu = \frac{\sum_{k=1}^N x_k}{N}$$

Στατιστική Ανάλυση Δεδομένων

Μονοδιάστατα Δεδομένα

3. Απόκλιση και Τυπική Απόκλιση

Μέσος Όρος

$$\mu = \frac{\sum_{k=1}^N x_k}{N}$$

Απόκλιση

$$\sigma^2 = \frac{\sum_{k=1}^N (x_k - \mu)^2}{N - 1}$$

Στατιστική Ανάλυση Δεδομένων

Μονοδιάστατα Δεδομένα

3. Απόκλιση και Τυπική Απόκλιση

Μέσος Όρος

$$\mu = \frac{\sum_{k=1}^N x_k}{N}$$

Απόκλιση

$$\sigma^2 = \frac{\sum_{k=1}^N (x_k - \mu)^2}{N - 1}$$

Τυπική Απόκλιση

$$\sigma = \sqrt{\frac{\sum_{k=1}^N (x_k - \mu)^2}{N - 1}}$$

Στατιστική Ανάλυση Δεδομένων

Μονοδιάστατα Δεδομένα

3. Απόκλιση και Τυπική Απόκλιση

Μέσος Όρος

$$\mu = \frac{\sum_{k=1}^N x_k}{N}$$

Απόκλιση

$$\sigma^2 = \frac{\sum_{k=1}^N (x_k - \mu)^2}{N - 1}$$

Τυπική Απόκλιση

$$\sigma = \sqrt{\frac{\sum_{k=1}^N (x_k - \mu)^2}{N - 1}}$$

Η Απόκλιση και η Τυπική
Απόκλιση είναι πάντα **θετικές**
ποσότητες

Στατιστική Ανάλυση Δεδομένων

Μονοδιάστατα Δεδομένα

3. Απόκλιση και Τυπική Απόκλιση

Μέσος Όρος

$$\mu = \frac{\sum_{k=1}^N x_k}{N}$$

Απόκλιση

$$\sigma^2 = \frac{\sum_{k=1}^N (x_k - \mu)^2}{N - 1}$$

Τυπική Απόκλιση

$$\sigma = \sqrt{\frac{\sum_{k=1}^N (x_k - \mu)^2}{N - 1}}$$

Η Απόκλιση και η Τυπική Απόκλιση είναι πάντα **θετικές ποσότητες**

	x
x_1	5
x_2	4
x_3	-3
x_4	7
x_5	9
x_6	-10

Στατιστική Ανάλυση Δεδομένων

Μονοδιάστατα Δεδομένα

3. Απόκλιση και Τυπική Απόκλιση

Μέσος Όρος

$$\mu = \frac{\sum_{k=1}^N x_k}{N}$$

Απόκλιση

$$\sigma^2 = \frac{\sum_{k=1}^N (x_k - \mu)^2}{N - 1}$$

Τυπική Απόκλιση

$$\sigma = \sqrt{\frac{\sum_{k=1}^N (x_k - \mu)^2}{N - 1}}$$

Η Απόκλιση και η Τυπική Απόκλιση είναι πάντα **θετικές ποσότητες**

	x
x_1	5
x_2	4
x_3	-3
x_4	7
x_5	9
x_6	-10

$$\mu = 2$$

Στατιστική Ανάλυση Δεδομένων

Μονοδιάστατα Δεδομένα

3. Απόκλιση και Τυπική Απόκλιση

Μέσος Όρος

$$\mu = \frac{\sum_{k=1}^N x_k}{N}$$

Απόκλιση

$$\sigma^2 = \frac{\sum_{k=1}^N (x_k - \mu)^2}{N - 1}$$

Τυπική Απόκλιση

$$\sigma = \sqrt{\frac{\sum_{k=1}^N (x_k - \mu)^2}{N - 1}}$$

Η Απόκλιση και η Τυπική Απόκλιση είναι πάντα **θετικές ποσότητες**

	x
x_1	5
x_2	4
x_3	-3
x_4	7
x_5	9
x_6	-10

$$\mu = 2$$

$$\sigma^2 = \frac{(5-2)^2 + (4-2)^2 + (-3-2)^2 + (7-2)^2 + (9-2)^2 + (-10-2)^2}{5} = 51.2$$

Στατιστική Ανάλυση Δεδομένων

Μονοδιάστατα Δεδομένα

3. Απόκλιση και Τυπική Απόκλιση

Μέσος Όρος

$$\mu = \frac{\sum_{k=1}^N x_k}{N}$$

Απόκλιση

$$\sigma^2 = \frac{\sum_{k=1}^N (x_k - \mu)^2}{N - 1}$$

Τυπική Απόκλιση

$$\sigma = \sqrt{\frac{\sum_{k=1}^N (x_k - \mu)^2}{N - 1}}$$

Η Απόκλιση και η Τυπική Απόκλιση είναι πάντα **θετικές ποσότητες**

	x
x_1	5
x_2	4
x_3	-3
x_4	7
x_5	9
x_6	-10

$$\mu = 2$$

$$\sigma^2 = \frac{(5-2)^2 + (4-2)^2 + (-3-2)^2 + (7-2)^2 + (9-2)^2 + (-10-2)^2}{5} = 51.2$$

$$\sigma = 7.1554$$

Στατιστική Ανάλυση Δεδομένων

Μονοδιάστατα Δεδομένα

3. Απόκλιση και Τυπική Απόκλιση

Μέσος Όρος

$$\mu = \frac{\sum_{k=1}^N x_k}{N}$$

Απόκλιση

$$\sigma^2 = \frac{\sum_{k=1}^N (x_k - \mu)^2}{N - 1}$$

Τυπική Απόκλιση

$$\sigma = \sqrt{\frac{\sum_{k=1}^N (x_k - \mu)^2}{N - 1}}$$

Η Απόκλιση και η Τυπική Απόκλιση είναι πάντα **θετικές ποσότητες**

	x
x_1	5
x_2	4
x_3	-3
x_4	7
x_5	9
x_6	-10

$$\mu = 2$$

$$\sigma^2 = \frac{(5-2)^2 + (4-2)^2 + (-3-2)^2 + (7-2)^2 + (9-2)^2 + (-10-2)^2}{5} = 51.2$$

$$\sigma = 7.1554$$



Στατιστική Ανάλυση Δεδομένων

Μονοδιάστατα Δεδομένα

3. Απόκλιση και Τυπική Απόκλιση

Μέσος Όρος

$$\mu = \frac{\sum_{k=1}^N x_k}{N}$$

Απόκλιση

$$\sigma^2 = \frac{\sum_{k=1}^N (x_k - \mu)^2}{N - 1}$$

Τυπική Απόκλιση

$$\sigma = \sqrt{\frac{\sum_{k=1}^N (x_k - \mu)^2}{N - 1}}$$

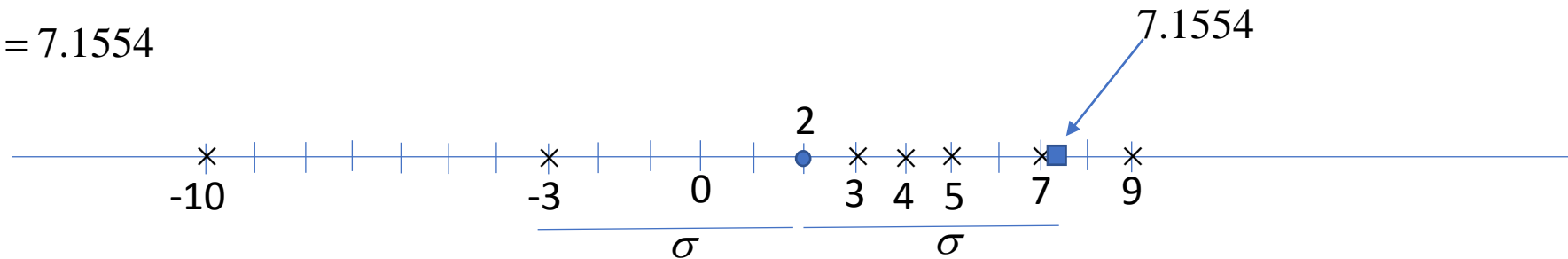
Η Απόκλιση και η Τυπική Απόκλιση είναι πάντα **θετικές ποσότητες**

	x
x_1	5
x_2	4
x_3	-3
x_4	7
x_5	9
x_6	-10

$$\mu = 2$$

$$\sigma^2 = \frac{(5-2)^2 + (4-2)^2 + (-3-2)^2 + (7-2)^2 + (9-2)^2 + (-10-2)^2}{5} = 51.2$$

$$\sigma = 7.1554$$



Στατιστική Ανάλυση Δεδομένων

Δυσδιάστατα Δεδομένα

1. Μέσος Όρος

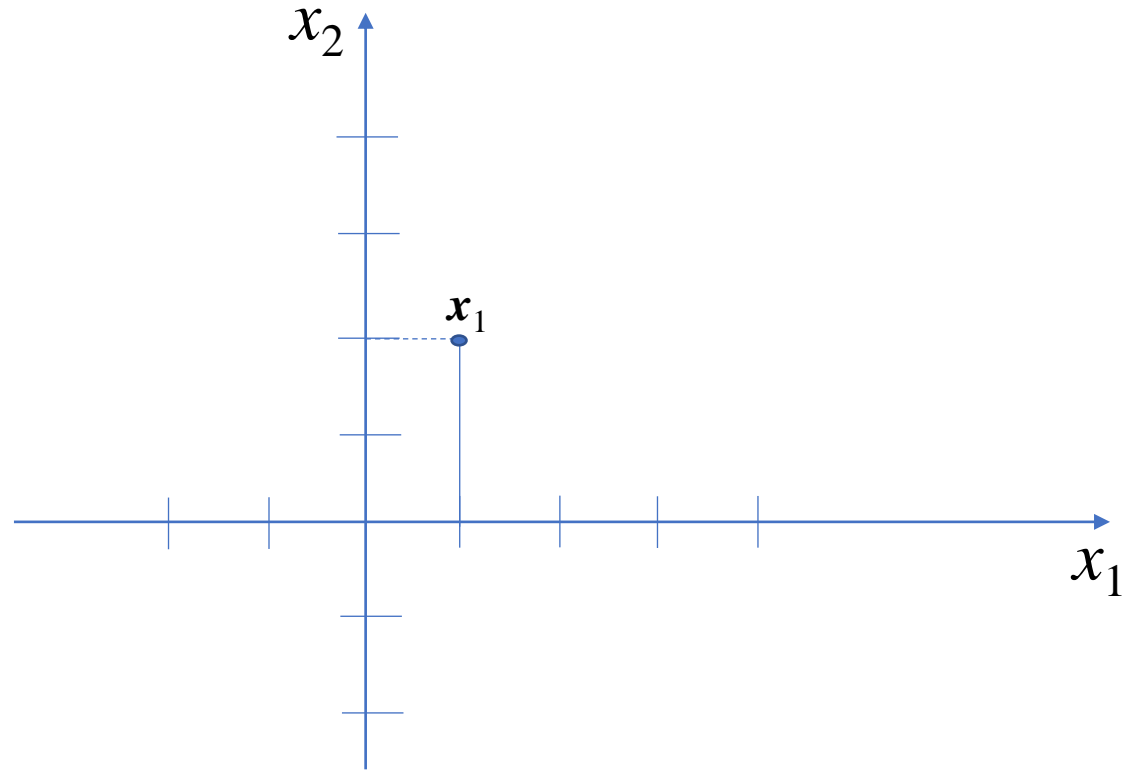
	x_1	x_2
x_1	1	2
x_2	4	2
x_3	2	-2
x_4	-1	4
x_5	4	-1

Στατιστική Ανάλυση Δεδομένων

Δυσδιάστατα Δεδομένα

1. Μέσος Όρος

	x_1	x_2
x_1	1	2
x_2	4	2
x_3	2	-2
x_4	-1	4
x_5	4	-1

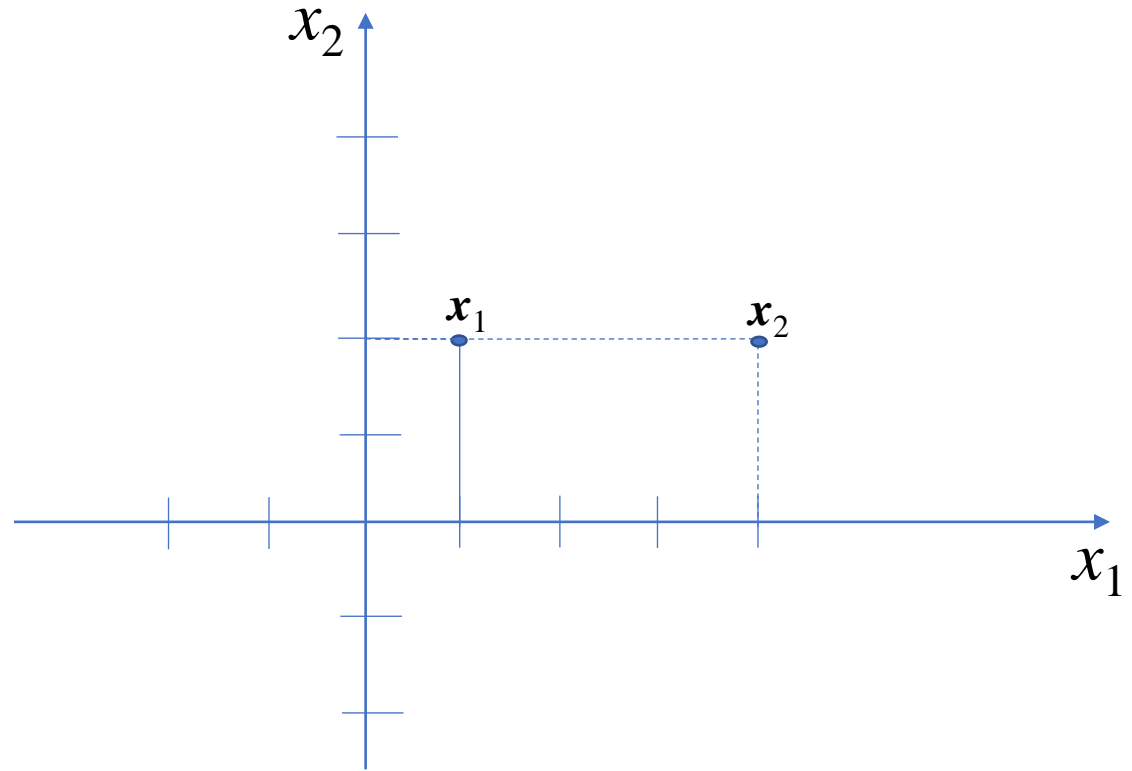


Στατιστική Ανάλυση Δεδομένων

Δυσδιάστατα Δεδομένα

1. Μέσος Όρος

	x_1	x_2
x_1	1	2
x_2	4	2
x_3	2	-2
x_4	-1	4
x_5	4	-1

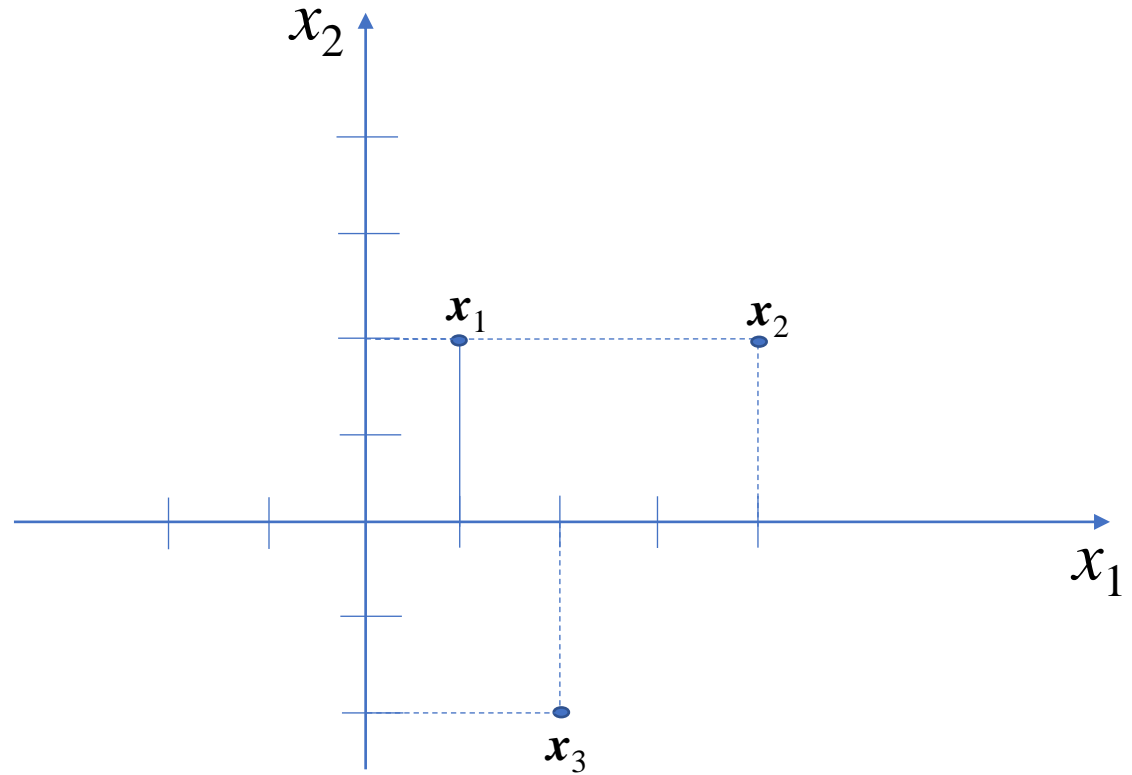


Στατιστική Ανάλυση Δεδομένων

Δυσδιάστατα Δεδομένα

1. Μέσος Όρος

	x_1	x_2
x_1	1	2
x_2	4	2
x_3	2	-2
x_4	-1	4
x_5	4	-1

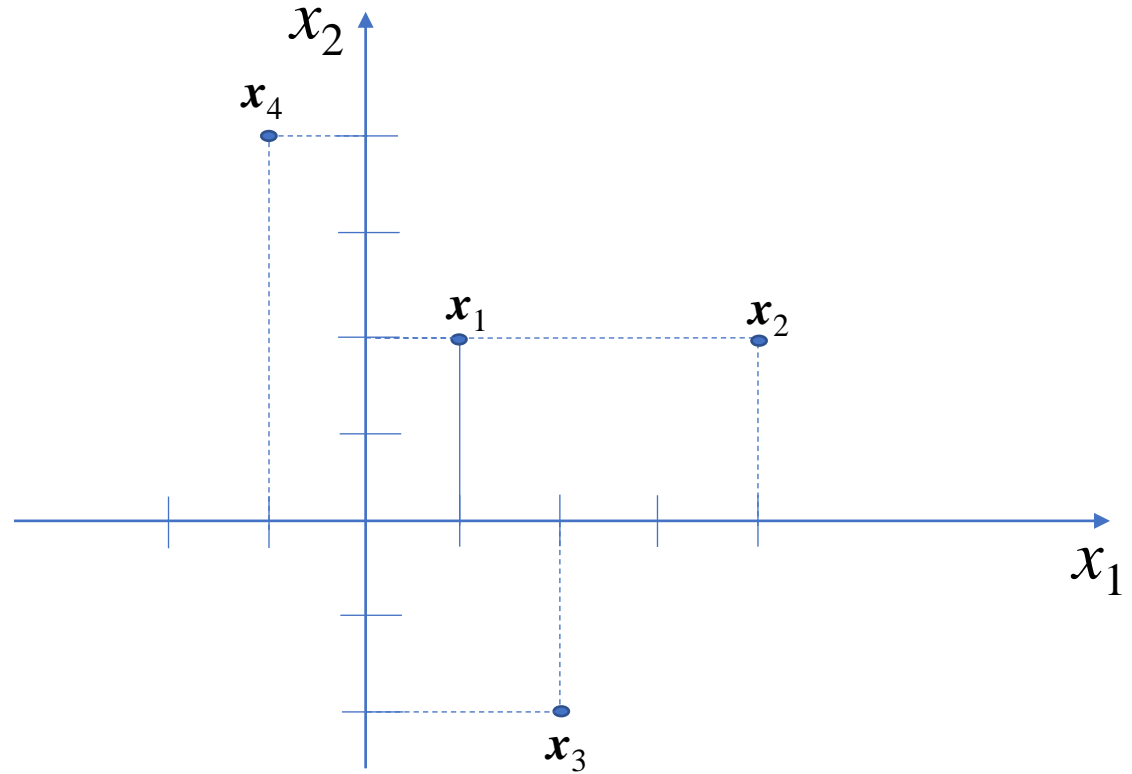


Στατιστική Ανάλυση Δεδομένων

Δυσδιάστατα Δεδομένα

1. Μέσος Όρος

	x_1	x_2
x_1	1	2
x_2	4	2
x_3	2	-2
x_4	-1	4
x_5	4	-1

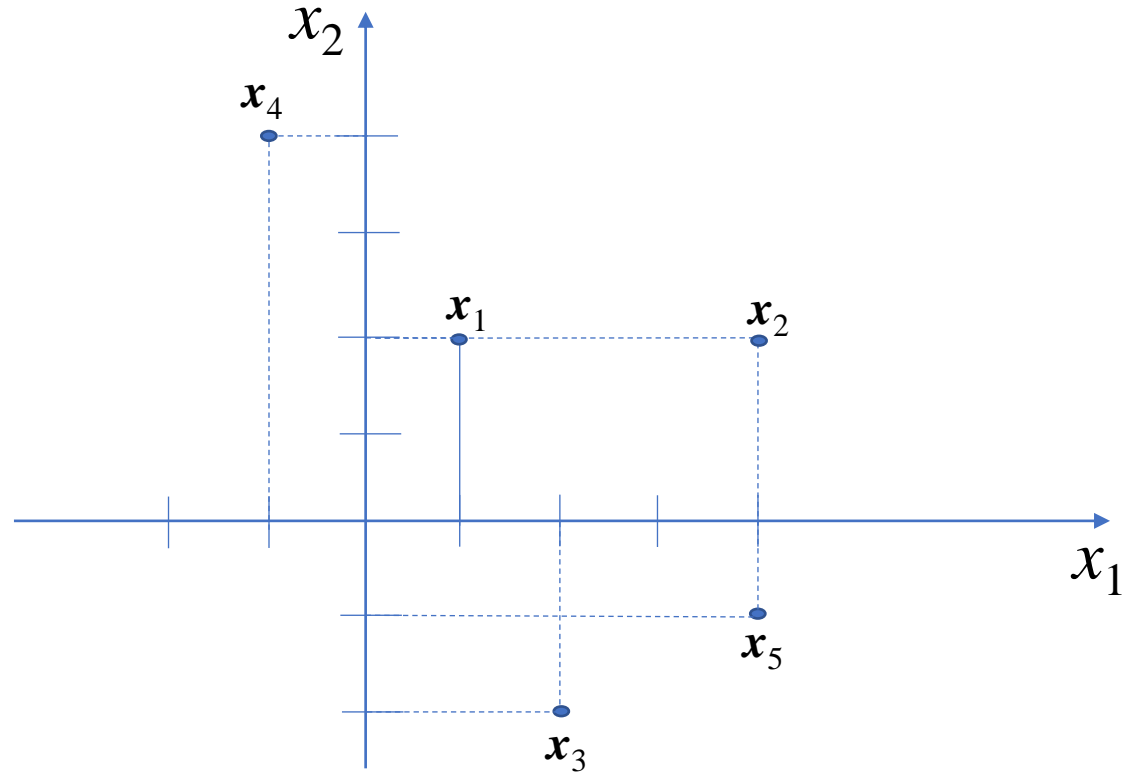


Στατιστική Ανάλυση Δεδομένων

Δυσδιάστατα Δεδομένα

1. Μέσος Όρος

	x_1	x_2
x_1	1	2
x_2	4	2
x_3	2	-2
x_4	-1	4
x_5	4	-1

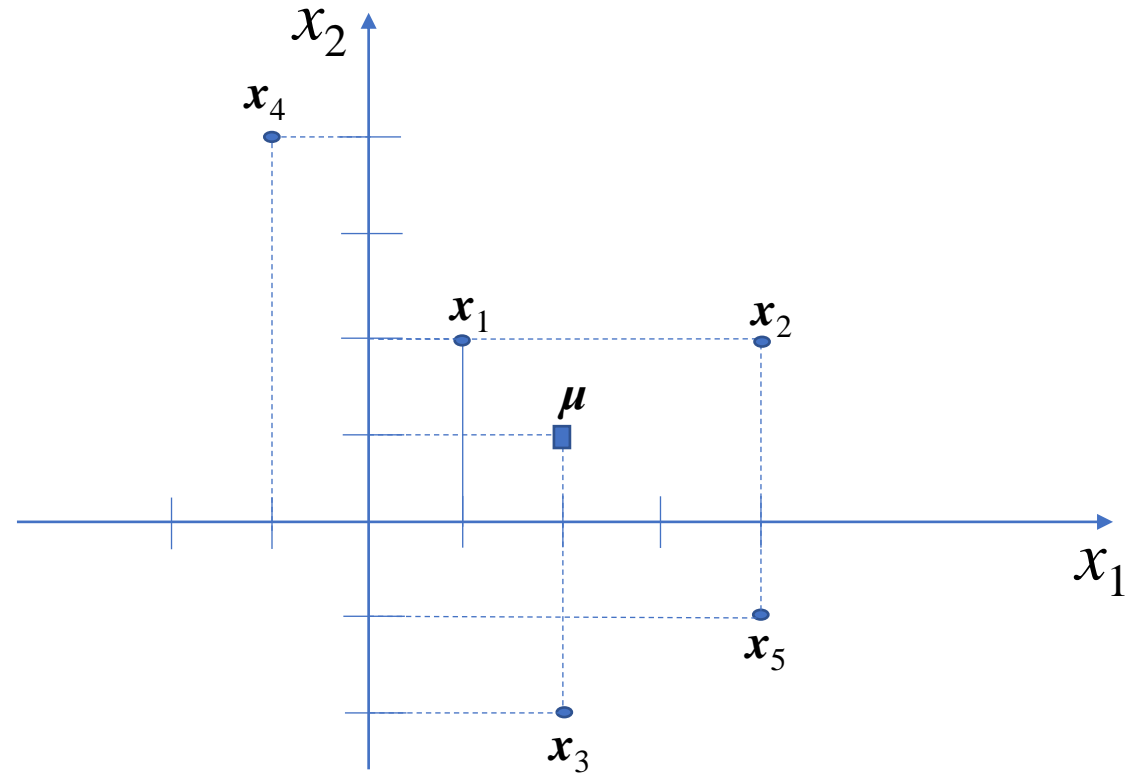


Στατιστική Ανάλυση Δεδομένων

Δυσδιάστατα Δεδομένα

1. Μέσος Όρος

	x_1	x_2
x_1	1	2
x_2	4	2
x_3	2	-2
x_4	-1	4
x_5	4	-1
μ	$\mu_1=2$	$\mu_2=1$

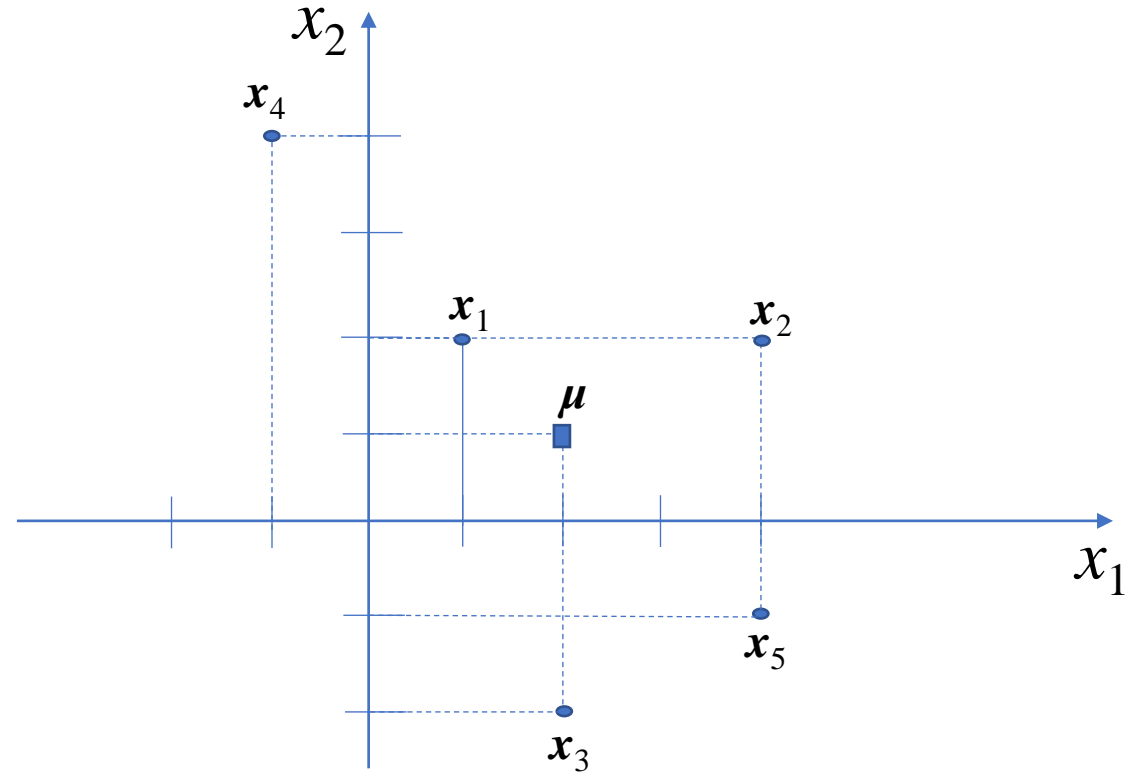


Στατιστική Ανάλυση Δεδομένων

Δυσδιάστατα Δεδομένα

1. Μέσος Όρος

	x_1	x_2
x_1	1	2
x_2	4	2
x_3	2	-2
x_4	-1	4
x_5	4	-1
μ	$\mu_1=2$	$\mu_2=1$



Ο μέσος όρος δυσδιάστατων διανυσματικών δεδομένων είναι και αυτός διανυσματικό διάνυσμα

Στατιστική Ανάλυση Δεδομένων

Πολυδιάστατα Δεδομένα

Μέσος Όρος

Πίνακας δεδομένων εισόδου: X

		x_1	x_2	...	x_j	...	x_p
x_1	1	x_{11}	x_{12}	...	x_{1j}	...	x_{1p}
x_2	2	x_{21}	x_{22}	...	x_{2j}	...	x_{2p}
...	...						
x_k	k	x_{k1}	x_{k2}	...	x_{kj}	...	x_{kp}
...	...						
x_N	N	x_{N1}	x_{N2}	...	x_{Nj}	...	x_{Np}
μ		μ_1	μ_2	...	μ_j	...	μ_p

Ο μέσος όρος πολυδιάστατων διανυσματικών δεδομένων είναι και αυτός διανυσματικό διάνυσμα

Συσταδοποίηση Δεδομένων

Συστάδα

Συλλογή αντικειμένων με κοινά χαρακτηριστικά

- Στο πλαίσιο των ΒΔ ερμηνεύεται ως μία ομάδα δεδομένων τα οποία είναι όμοια μεταξύ τους

Συσταδοποίηση Δεδομένων

Συστάδα

Συλλογή αντικειμένων με κοινά χαρακτηριστικά

- Στο πλαίσιο των ΒΔ ερμηνεύεται ως μία ομάδα δεδομένων τα οποία είναι όμοια μεταξύ τους
- Η ομοιότητα αναφέρεται στο ότι οι Ευκλείδειες αποστάσεις μεταξύ των δεδομένων είναι μικρές με αποτέλεσμα τα δεδομένα να είναι πολύ κοντά το ένα με το άλλο

Συσταδοποίηση Δεδομένων

Συστάδα

Συλλογή αντικειμένων με κοινά χαρακτηριστικά

- Στο πλαίσιο των ΒΔ ερμηνεύεται ως μία ομάδα δεδομένων τα οποία είναι όμοια μεταξύ τους
- Η ομοιότητα αναφέρεται στο ότι οι Ευκλείδειες αποστάσεις μεταξύ των δεδομένων είναι μικρές με αποτέλεσμα τα δεδομένα να είναι πολύ κοντά το ένα με το άλλο

Συσταδοποίηση

Ας υποθέσουμε ότι έχουμε ένα σύνολο δεδομένων (πολυδιάστατων ώστε να “πιάσουμε” την γενική περίπτωση)

- Με τον όρο συσταδοποίηση εννοούμε τον διαμερισμό του συνόλου σε έναν αριθμό συστάδων (ομάδων) με τις παρακάτω ιδιότητες:

Συσταδοποίηση Δεδομένων

Συστάδα

Συλλογή αντικειμένων με κοινά χαρακτηριστικά

- Στο πλαίσιο των ΒΔ ερμηνεύεται ως μία ομάδα δεδομένων τα οποία είναι όμοια μεταξύ τους
- Η ομοιότητα αναφέρεται στο ότι οι Ευκλείδειες αποστάσεις μεταξύ των δεδομένων είναι μικρές με αποτέλεσμα τα δεδομένα να είναι πολύ κοντά το ένα με το άλλο

Συσταδοποίηση

Ας υποθέσουμε ότι έχουμε ένα σύνολο δεδομένων (πολυδιάστατων ώστε να “πιάσουμε” την γενική περίπτωση)

- Με τον όρο συσταδοποίηση εννοούμε τον διαμερισμό του συνόλου σε έναν αριθμό συστάδων (ομάδων) με τις παρακάτω ιδιότητες:
 - Συμπαγότητα της κάθε ομάδας: Δεδομένα που ανήκουν στην ίδια συστάδα να είναι όσο το δυνατόν πιο όμοια μεταξύ τους

Συσταδοποίηση Δεδομένων

Συστάδα

Συλλογή αντικειμένων με κοινά χαρακτηριστικά

- Στο πλαίσιο των ΒΔ ερμηνεύεται ως μία ομάδα δεδομένων τα οποία είναι όμοια μεταξύ τους
- Η ομοιότητα αναφέρεται στο ότι οι Ευκλείδειες αποστάσεις μεταξύ των δεδομένων είναι μικρές με αποτέλεσμα τα δεδομένα να είναι πολύ κοντά το ένα με το άλλο

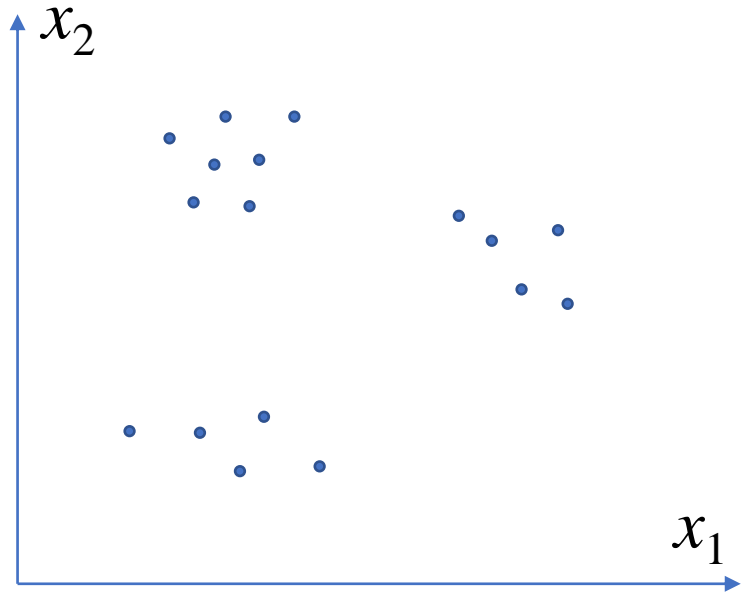
Συσταδοποίηση

Ας υποθέσουμε ότι έχουμε ένα σύνολο δεδομένων (πολυδιάστατων ώστε να “πιάσουμε” την γενική περίπτωση)

- Με τον όρο συσταδοποίηση εννοούμε τον διαμερισμό του συνόλου σε έναν αριθμό συστάδων (ομάδων) με τις παρακάτω ιδιότητες:
 - Συμπαγότητα της κάθε ομάδας: Δεδομένα που ανήκουν στην ίδια συστάδα να είναι όσο το δυνατόν πιο όμοια μεταξύ τους
 - Διαχωρισσιμότητα μεταξύ ομάδων: Δεδομένα που ανήκουν σε διαφορετικές συστάδες να είναι όσο το δυνατόν πιο ανόμοια

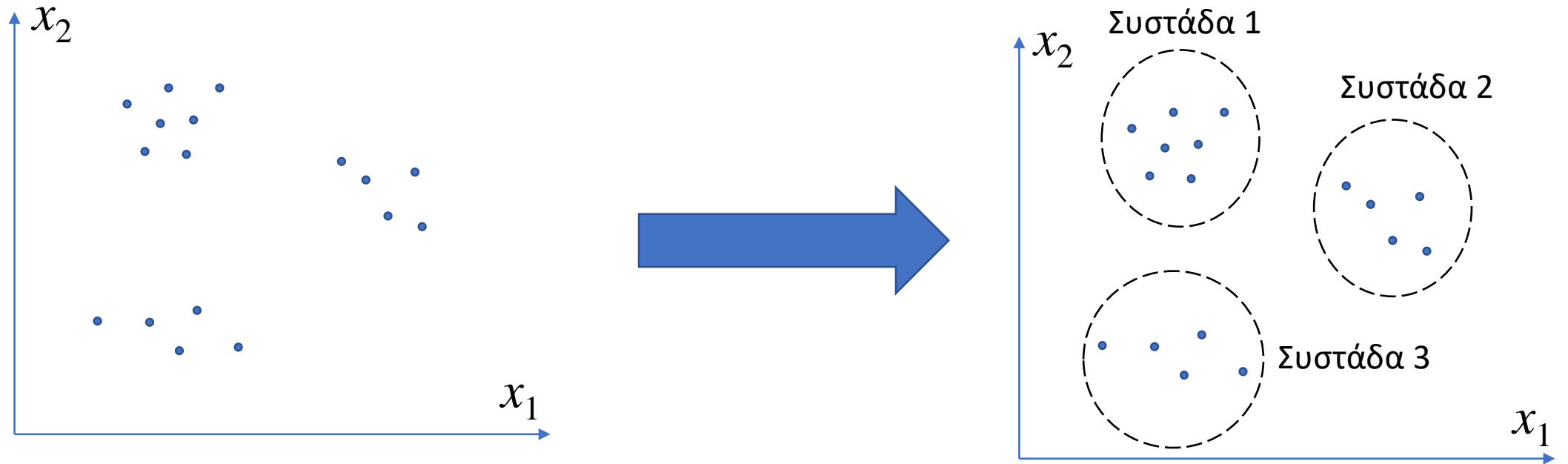
Συσταδοποίηση Δεδομένων

Διαχωρίσιμες και Συμπαγείς Συστάδες



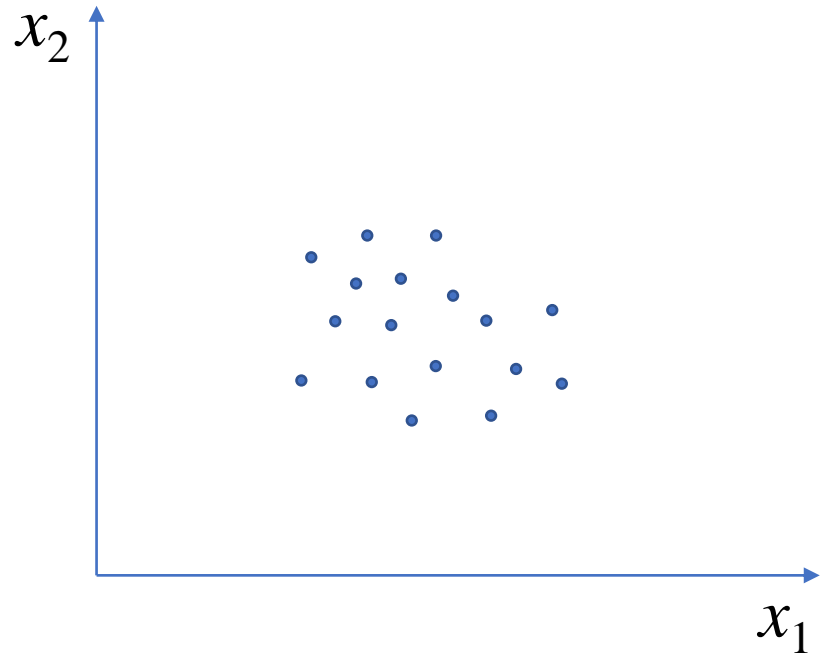
Συσταδοποίηση Δεδομένων

Διαχωρίσιμες και Συμπαγείς Συστάδες



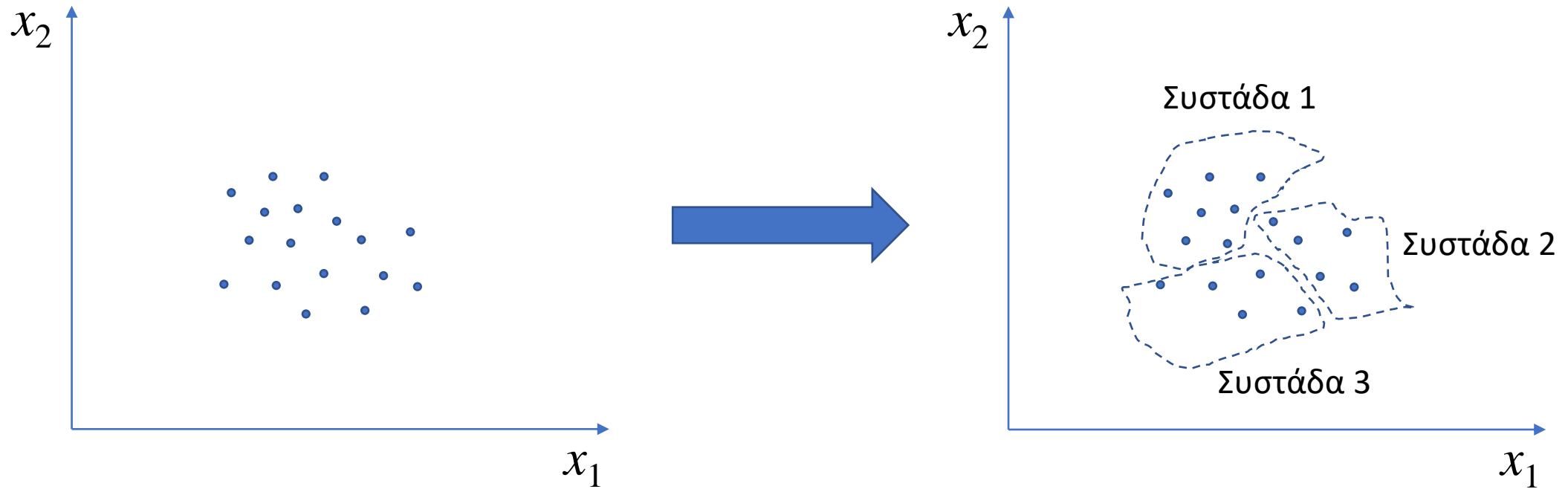
Συσταδοποίηση Δεδομένων

Μη Διαχωρίσιμες Συστάδες



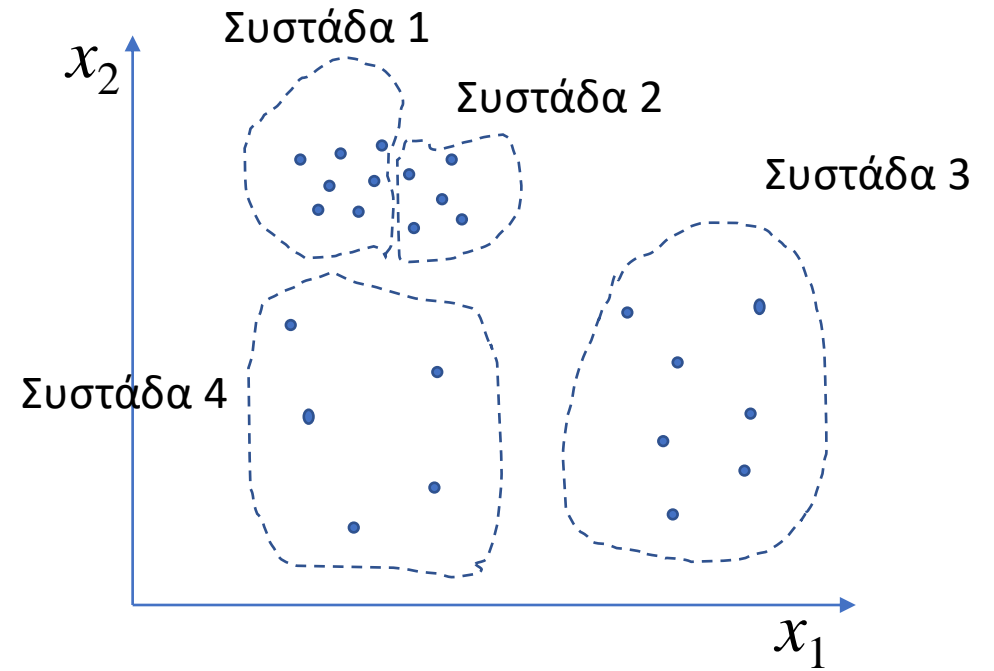
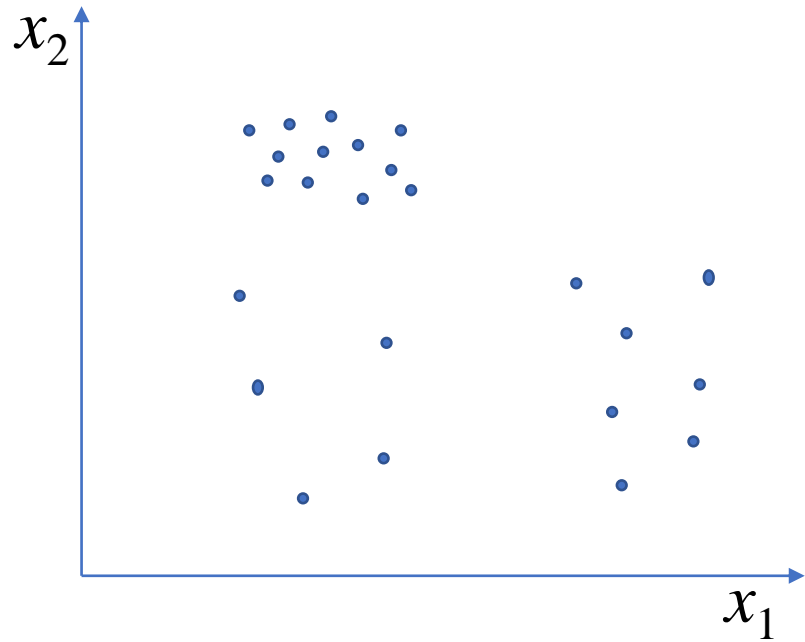
Συσταδοποίηση Δεδομένων

Μη Διαχωρίσιμες Συστάδες



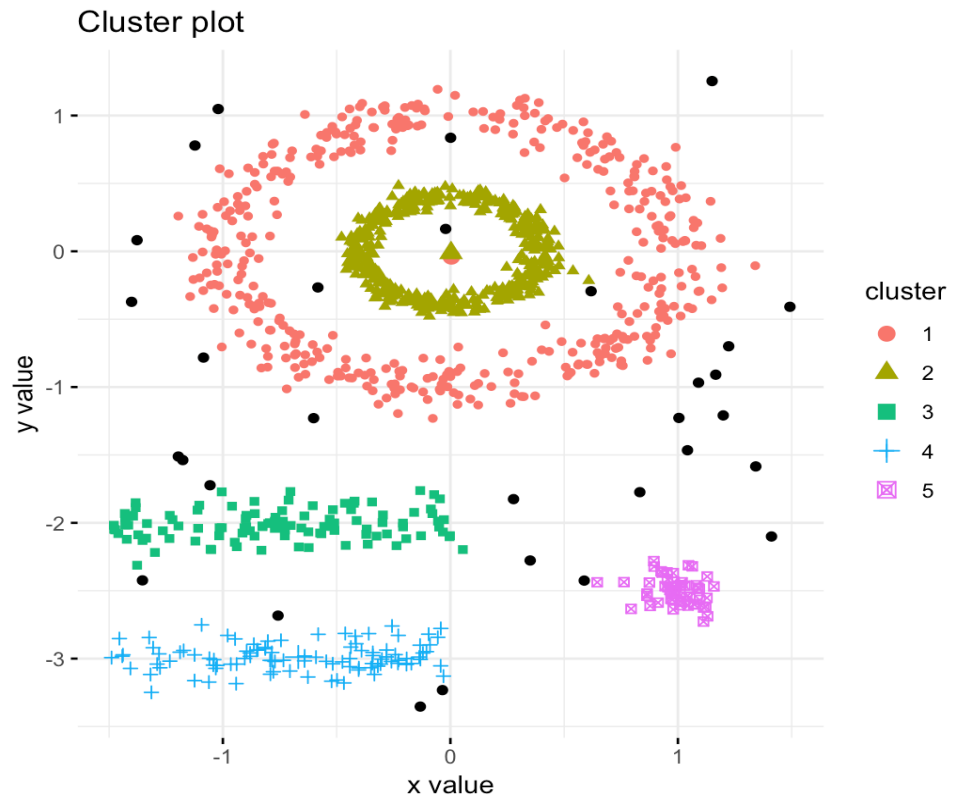
Συσταδοποίηση Δεδομένων

Συμπαγείς και Μη Συμπαγής Συστάδες



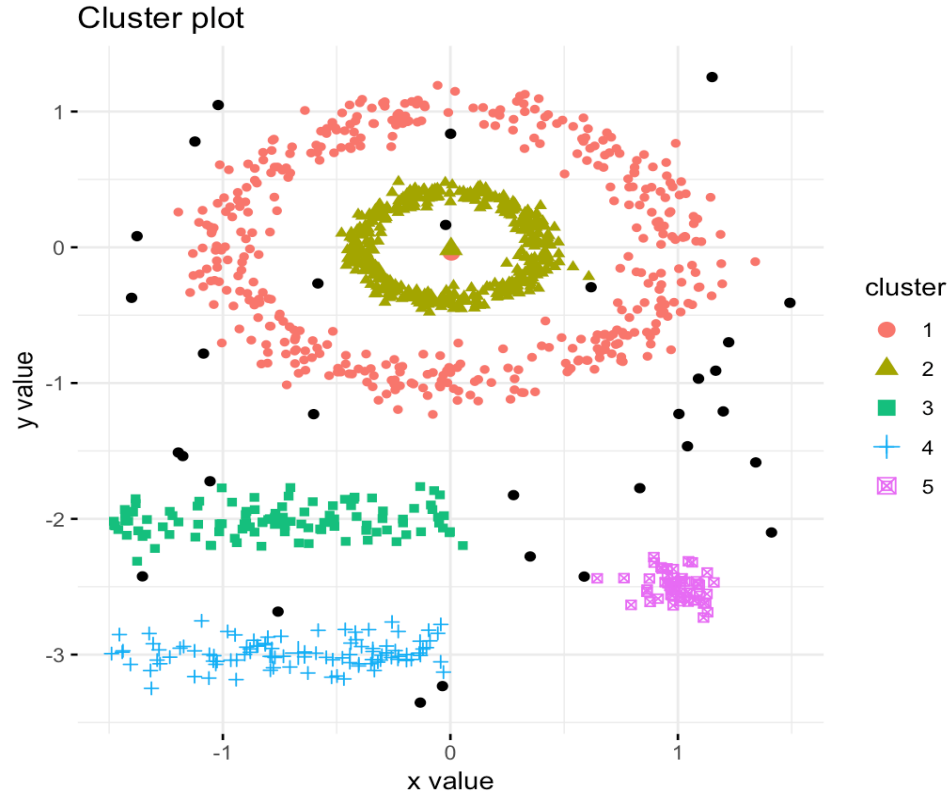
Συσταδοποίηση Δεδομένων

Συστάδες Πραγματικών Δεδομένων

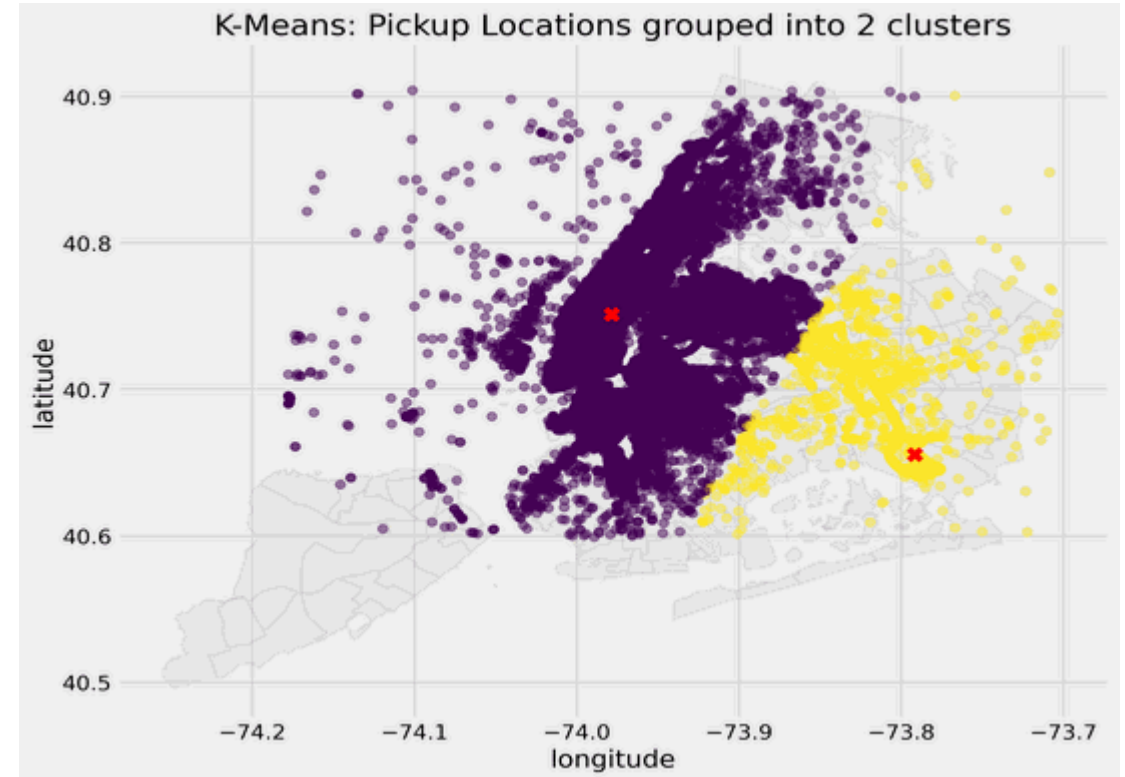


Συσταδοποίηση Δεδομένων

Συστάδες Πραγματικών Δεδομένων



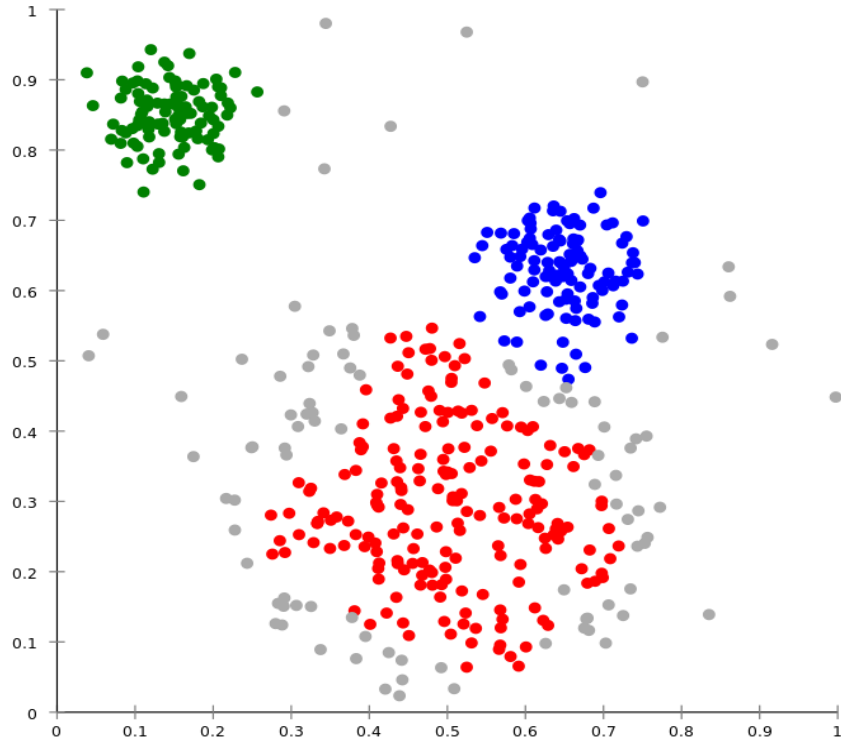
datanovia.com



towardsdatascience.com

Συσταδοποίηση Δεδομένων

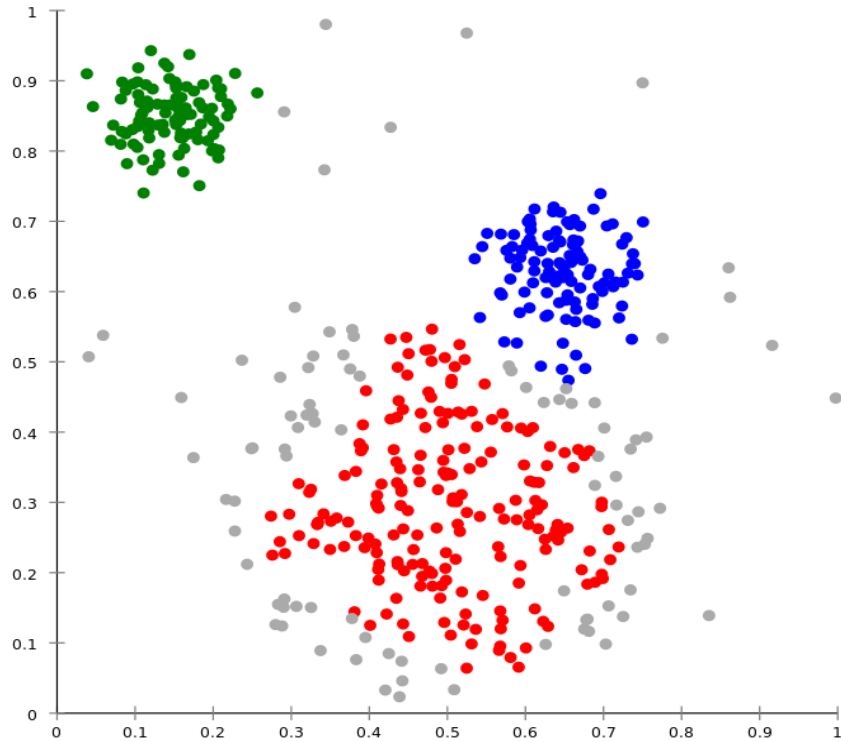
Συστάδες Πραγματικών Δεδομένων



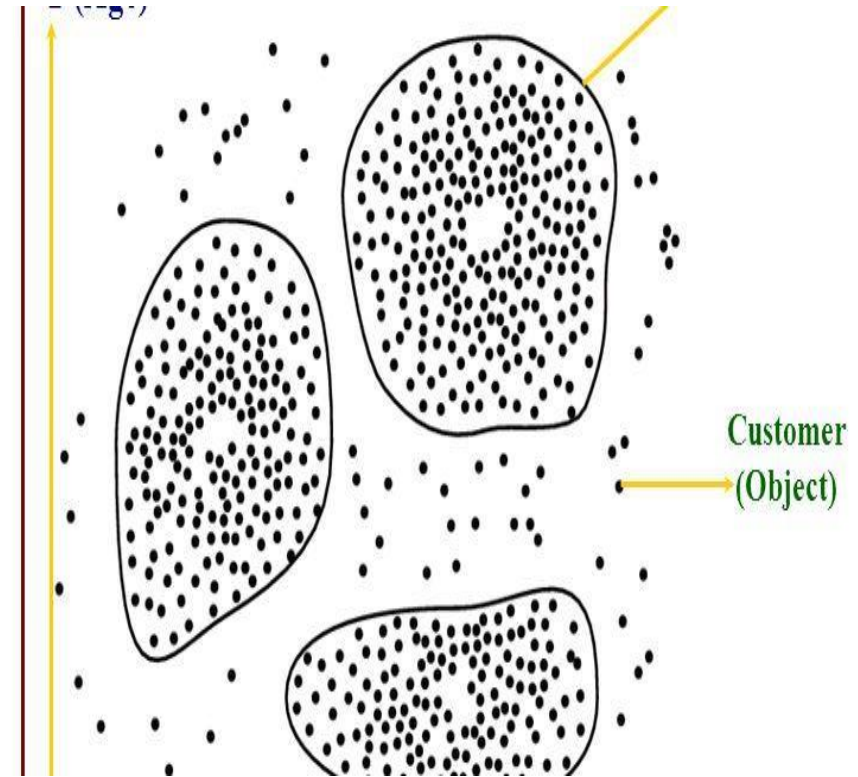
wikipedia

Συσταδοποίηση Δεδομένων

Συστάδες Πραγματικών Δεδομένων



wikipedia



<https://analyticsbuddhu.wordpress.com>

Συσταδοποίηση Δεδομένων

Παράγοντες που επηρεάζουν μία Συσταδοποίηση

- Το μέγεθος των δεδομένων
- Ο αλγόριθμος συσταδοποίησης που χρησιμοποιείται

Συσταδοποίηση Δεδομένων

Παράγοντες που επηρεάζουν μία Συσταδοποίηση

- Το μέγεθος των δεδομένων
- Ο αλγόριθμος συσταδοποίησης που χρησιμοποιείται

Χαρακτηριστικά των Αλγόριθμων Συσταδοποίησης

- Ανήκουν στην κατηγορία της Μη-Επιβλεπόμενης Μάθησης (Unsupervised Learning)
- Είναι επαναληπτικοί αλγόριθμοι
- Οι αρχικές τιμές των παραμέτρων τους μπορεί να επιλεγθούν τυχαία ή από τα ίδια τα δεδομένα

Συσταδοποίηση Δεδομένων

Είδη Αλγόριθμων Συσταδοποίησης

- Αλγόριθμοι Κατηγοριοποίησης (Batch Clustering Algorithms)

Συσταδοποίηση Δεδομένων

Είδη Αλγόριθμων Συσταδοποίησης

- **Αλγόριθμοι Κατηγοριοποίησης (Batch Clustering Algorithms)**
Επεξεργάζονται ταυτόχρονα τα δεδομένα και μετά το τέλος της επαναληπτικής διαδικασίας προκύπτουν οι συστάδες

Συσταδοποίηση Δεδομένων

Είδη Αλγόριθμων Συσταδοποίησης

- **Αλγόριθμοι Κατηγοριοποίησης (Batch Clustering Algorithms)**
Επεξεργάζονται ταυτόχρονα τα δεδομένα και μετά το τέλος της επαναληπτικής διαδικασίας προκύπτουν οι συστάδες
- Χαρακτηριστικά παραδείγματα τέτοιων αλγόριθμων είναι τα παρακάτω:
 - k-means
 - fuzzy k-means, κλπ

Συσταδοποίηση Δεδομένων

Είδη Αλγόριθμων Συσταδοποίησης

- **Αλγόριθμοι Κατηγοριοποίησης (Batch Clustering Algorithms)**
Επεξεργάζονται ταυτόχρονα τα δεδομένα και μετά το τέλος της επαναληπτικής διαδικασίας προκύπτουν οι συστάδες
- Χαρακτηριστικά παραδείγματα τέτοιων αλγόριθμων είναι τα παρακάτω:
 - k-means
 - fuzzy k-means, κλπ
- **Ιεραρχικοί Αλγόριθμοι (Hierarchical Algorithms)**
Είναι κυρίως σειριακοί με την έννοια ότι εξετάζουν το κάθε δεδομένο ξεχωριστά

Συσταδοποίηση Δεδομένων

Είδη Αλγόριθμων Συσταδοποίησης

- **Αλγόριθμοι Κατηγοριοποίησης (Batch Clustering Algorithms)**

Επεξεργάζονται ταυτόχρονα τα δεδομένα και μετά το τέλος της επαναληπτικής διαδικασίας προκύπτουν οι συστάδες

- Χαρακτηριστικά παραδείγματα τέτοιων αλγόριθμων είναι τα παρακάτω:

- k-means
- fuzzy k-means, κλπ

- **Ιεραρχικοί Αλγόριθμοι (Hierarchical Algorithms)**

Είναι κυρίως σειριακοί με την έννοια ότι εξετάζουν το κάθε δεδομένο ξεχωριστά

- Συσσωρευτικοί (Agglomerative Clustering Algorithms)

Ένας Συσσωρευτικός Αλγόριθμος θεωρεί αρχικά ότι κάθε δεδομένα είναι μία συστάδες. Σε κάθε επόμενο βήμα αν εντοπίζει δύο όμοιες (πολύ κοντινές) συστάδες τις ενώνει σε μία. Σε κάθε βήμα μπορεί να ενώσει πολλά ζεύγη συστάδων.

Η διαδικασία τελειώνει όταν σχηματιστεί ο πιο αποδοτικός αριθμός συστάδων, οι οποίες περιγράφουν καλύτερα τα δεδομένα

Συσταδοποίηση Δεδομένων

Είδη Αλγόριθμων Συσταδοποίησης

- **Αλγόριθμοι Κατηγοριοποίησης (Batch Clustering Algorithms)**

Επεξεργάζονται ταυτόχρονα τα δεδομένα και μετά το τέλος της επαναληπτικής διαδικασίας προκύπτουν οι συστάδες

- Χαρακτηριστικά παραδείγματα τέτοιων αλγόριθμων είναι τα παρακάτω:

- k-means
- fuzzy k-means, κλπ

- **Ιεραρχικοί Αλγόριθμοι (Hierarchical Algorithms)**

Είναι κυρίως σειριακοί με την έννοια ότι εξετάζουν το κάθε δεδομένο ξεχωριστά

- Συσσωρευτικοί (Agglomerative Clustering Algorithms)

Ένας Συσσωρευτικός Αλγόριθμος θεωρεί αρχικά ότι κάθε δεδομένα είναι μία συστάδες. Σε κάθε επόμενο βήμα αν εντοπίζει δύο όμοιες (πολύ κοντινές) συστάδες τις ενώνει σε μία. Σε κάθε βήμα μπορεί να ενώσει πολλά ζεύγη συστάδων.

Η διαδικασία τελειώνει όταν σχηματιστεί ο πιο αποδοτικός αριθμός συστάδων, οι οποίες περιγράφουν καλύτερα τα δεδομένα

- Διαιρετικοί (Divisive Clustering Algorithms)

Ένας Διαιρετικός Αλγόριθμος Συσταδοποίησης αρχικά θεωρεί ότι όλα τα δεδομένα είναι μία συστάδα. Στην συνέχεια, σε κάθε επανάληψη διαχωρίζει την αρχική σε μικρότερες και αυτές που προκύπτουν σε ακόμα μικρότερες.

Η διαδικασία τελειώνει όταν σχηματιστεί ο πιο αποδοτικός αριθμός συστάδων, οι οποίες περιγράφουν καλύτερα τα δεδομένα

ΚΑΛΟ ΑΠΟΓΕΥΜΑ