



# ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ *«Διοίκηση στη Ναυτική Επιστήμη και Τεχνολογία»*

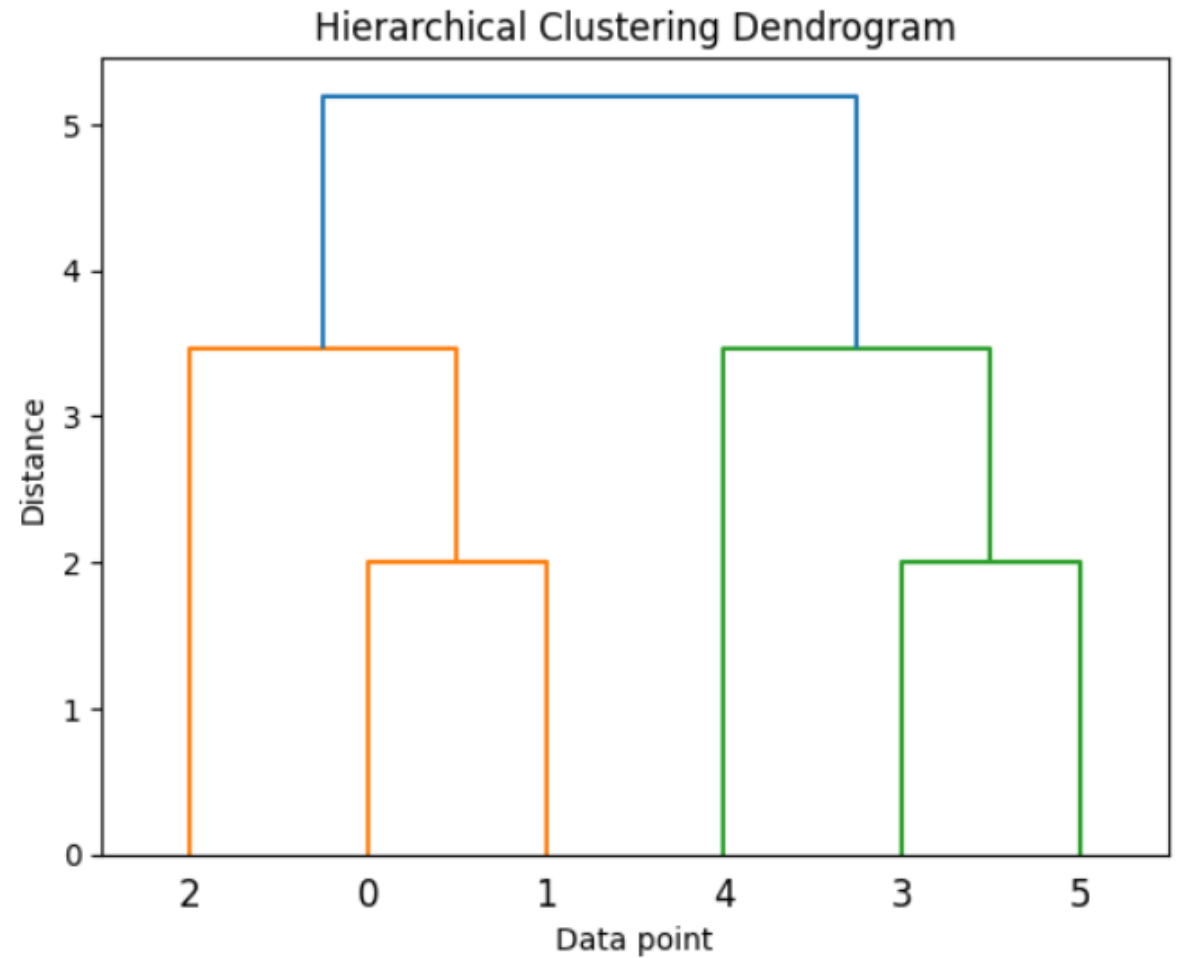


*Image by David Mark from Pixabay*

**Διαχείριση μεγάλου όγκου δεδομένων στη Ναυτιλία και το Θαλάσσιο Περιβάλλον**

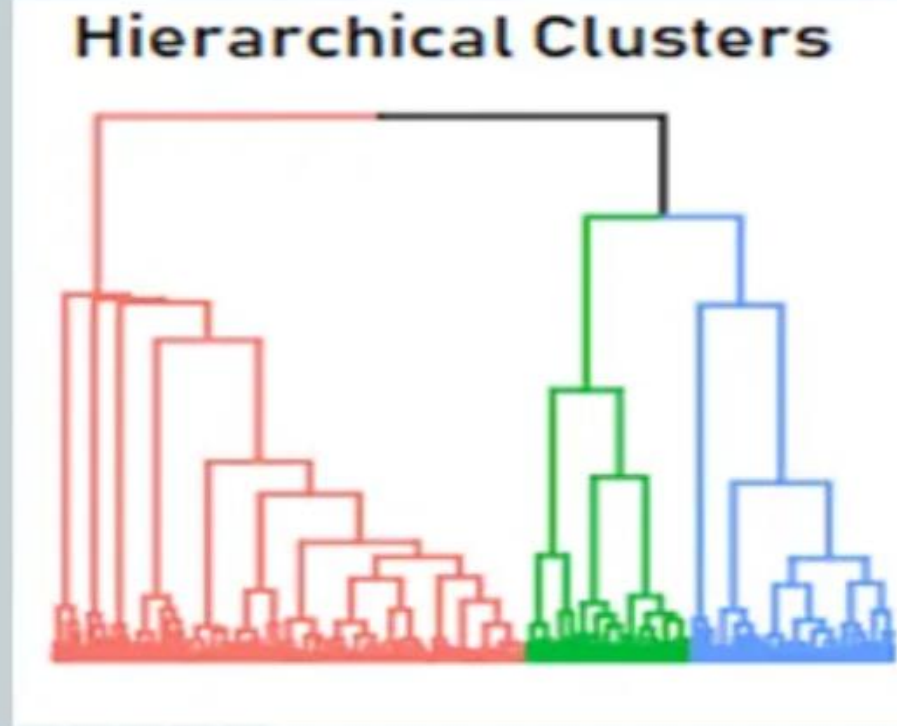
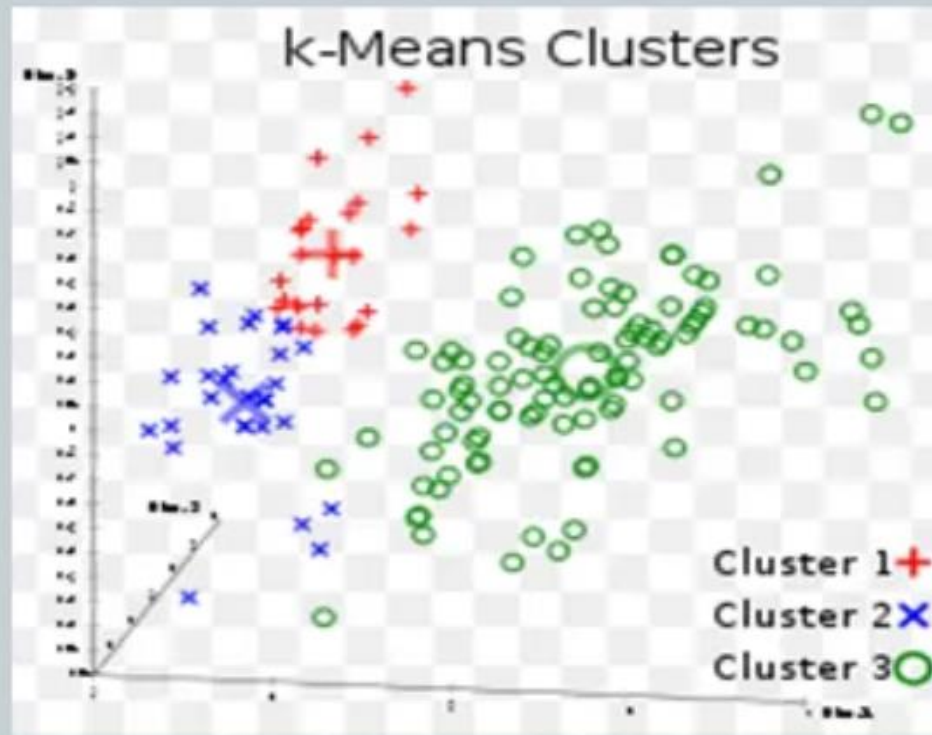
**Διδάσκοντες: Γ. Γαλάνης, Δ. Κουλουμπού**

# Ιεραρχικοί Αλγόριθμοι Συσταδοποίησης



*Hierarchical Clustering Dendrogram*

# Ιεραρχικοί Αλγόριθμοι Συσταδοποίησης

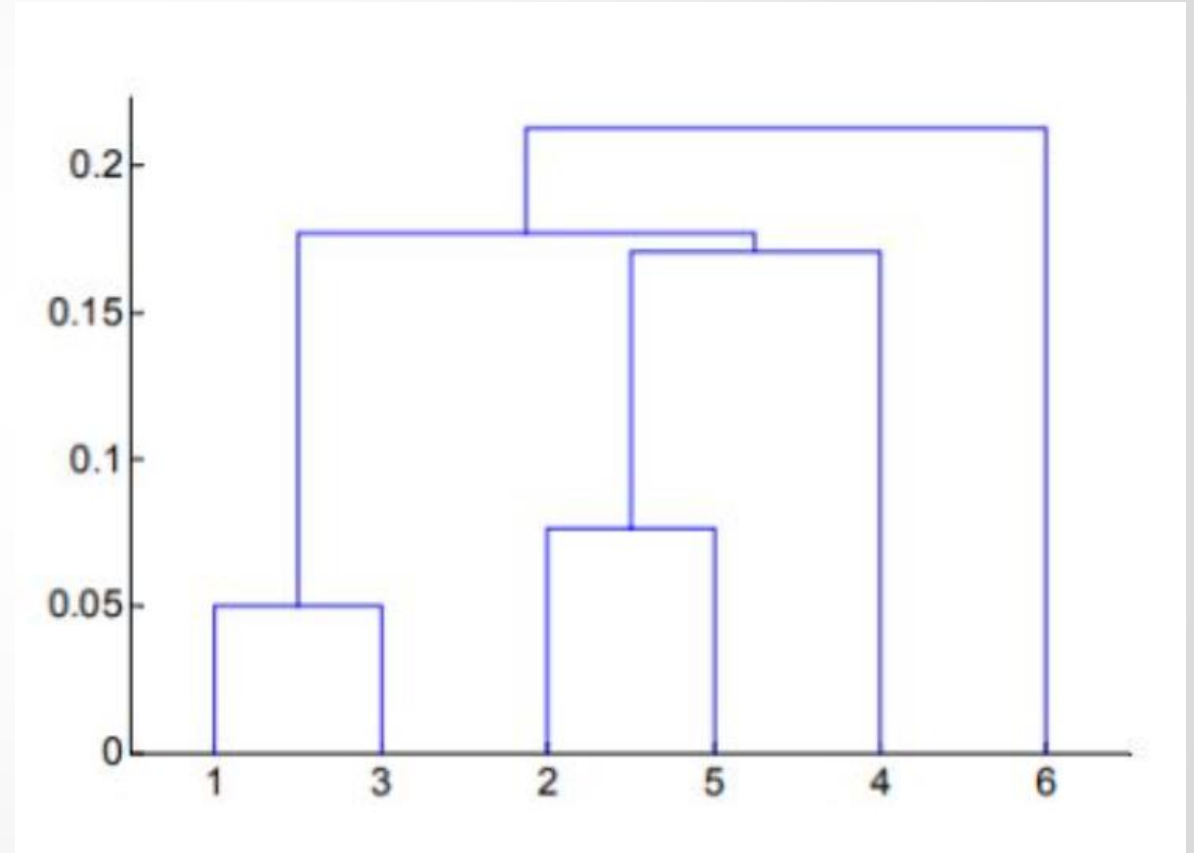


# Συσταδοποίηση – Clustering

- Όπως ήδη έχουμε αναφέρει στο **πρόβλημα της συσταδοποίησης** μας δίνεται ένα σύνολο  $n$  παρατηρήσεων, και για κάθε παρατήρηση έχουμε δεδομένα από  $N$  μεταβλητές  $X_1, X_2, \dots, X_N$ .

# Ιεραρχική Συσταδοποίηση

- Η ιεραρχική συσταδοποίηση παράγει ένα σύνολο από συστάδες (clusters) σε ένα ιεραρχικό δένδρο.



# Ιεραρχική Συσταδοποίηση

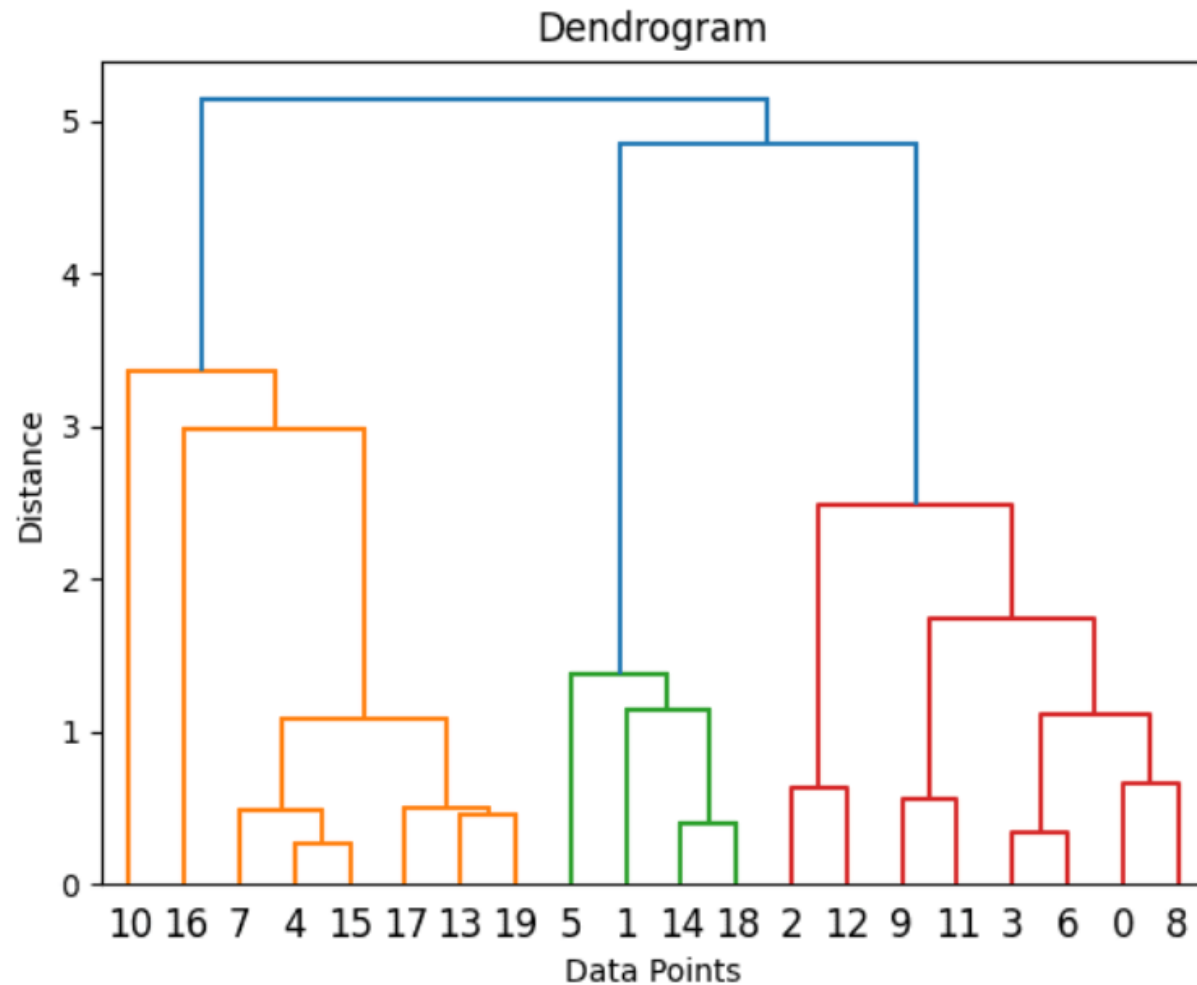
- Το δενδροδιάγραμμα παρουσιάζει τη σχέση μεταξύ των στοιχείων του δείγματος.
- Επίσης καταδεικνύει πώς τα στοιχεία έχουν ομαδοποιηθεί στα διαφορετικά βήματα της διαδικασία συσταδοποίησης.



# Ιεραρχικοί Αλγόριθμοι Συσταδοποίησης

- Οι ιεραρχικοί αλγόριθμοι μπορούν να αναπαρασταθούν πλήρως με δενδρογράμματα, δηλαδή με δενδρικά διαγράμματα, τα οποία παρουσιάζουν τη διάταξη των συστάδων που δημιουργήθηκαν από την ιεραρχική συσταδοποίηση.

# Ιεραρχικοί Αλγόριθμοι Συσταδοποίησης



# Συσταδοποίηση – Clustering

- Αρχικά, το σύνολο των δεδομένων δίνεται χωρίς τις αντίστοιχες κλάσεις ή ετικέτες και χρειαζόμαστε κάποιον αλγόριθμο, ο οποίος θα ομαδοποιήσει αυτόματα τα δεδομένα σε συστάδες.



# Συσταδοποίηση – Clustering

- Ένα από τα βασικά πλεονεκτήματα της ιεραρχικής συσταδοποίησης είναι ότι δεν χρειάζεται να ορίσουμε έναν συγκεκριμένο αριθμό συστάδων.
- Οποιοσδήποτε αριθμός μπορεί να επιτευχθεί, κόβοντας το δένδροδιάγραμμα σε ένα κατάλληλο επίπεδο.

# Συσταδοποίηση – Clustering

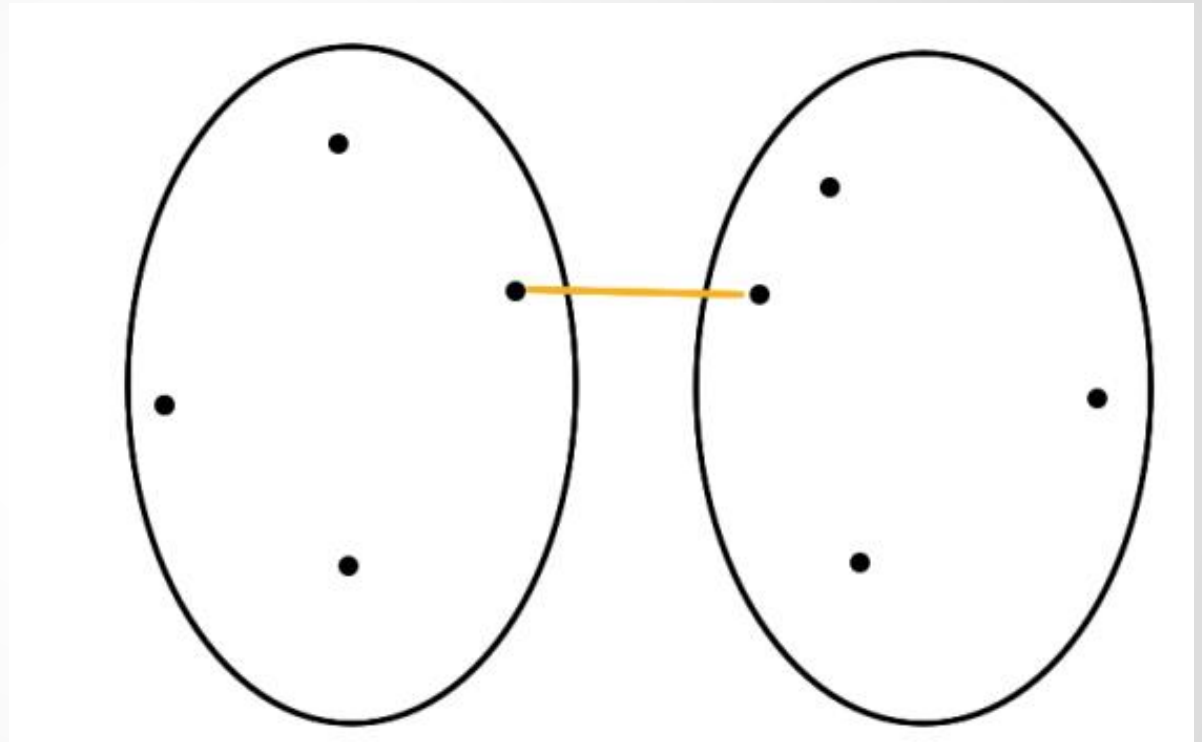
- Στη διαδικασία της ομαδοποίησης των συστάδων σημαντικό ρόλο έχει η απόσταση μεταξύ των συστάδων
- Η **απόσταση μεταξύ δύο συστάδων** μπορεί να οριστεί με έναν από τους εξής διαφορετικούς τρόπους:

# Απόσταση Συστάδων – Ελάχιστη Απόσταση

- **Ελάχιστη Απόσταση:** Είναι η μικρότερη απόσταση μεταξύ δύο σημείων των δύο συστάδων.

# Απόσταση Συστάδων

Minimum Απόσταση, ως μέτρο  
απόστασης δύο συστάδων



# Απόσταση Συστάδων – Ελάχιστη Απόσταση

- Η ελάχιστη απόσταση ως απόσταση δύο συστάδων βασίζεται στα δύο πιο όμοια (πιο γειτονικά) σημεία στις διαφορετικές συστάδες.
- Καθορίζεται από ένα ζεύγος τιμών, δηλαδή μια ακμή (link) του γραφήματος γειτνίασης.
- Η μέθοδος αυτή ονομάζεται και **μέθοδος συσταδοποίησης κοντινότερου γείτονα ή απλού συνδέσμου (single link)**.

# Απόσταση Συστάδων – Ελάχιστη Απόσταση

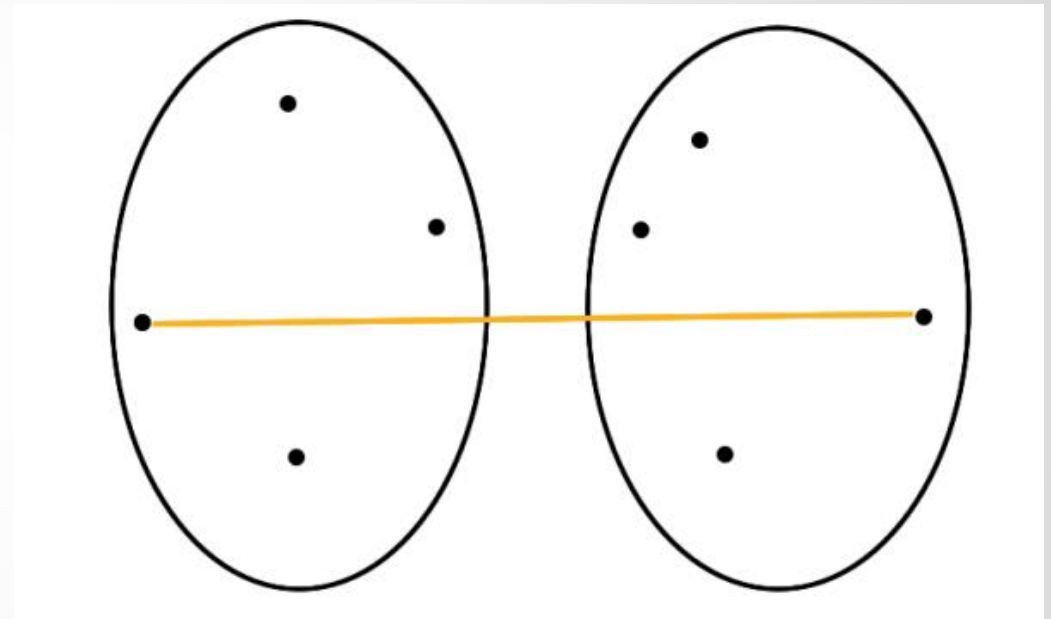
- Τα προτερήματα αυτής της μεθόδου είναι ότι δημιουργούνται συνεχόμενες συστάδες, ενώ μπορεί να χειριστεί μη ελλειπτικά σχήματα.
- Το βασικό μειονέκτημα είναι η ευαισθησία στον θόρυβο και στις ακραίες τιμές (outliers).

# Απόσταση Συστάδων – Μέγιστη Απόσταση

- **Μέγιστη Απόσταση:** Είναι η μεγαλύτερη απόσταση μεταξύ δύο σημείων των δύο συστάδων.

# Απόσταση Συστάδων

Maximum Απόσταση, ως μέτρο απόστασης δύο  
συστάδων



# Απόσταση Συστάδων – Μέγιστη Απόσταση

- Η μέγιστη απόσταση ως απόσταση δύο συστάδων βασίζεται στα δυο λιγότερο όμοια (πιο απομακρυσμένα) σημεία στις διαφορετικές συστάδες.
- Καθορίζεται από ένα ζεύγος τιμών, δηλαδή μια ακμή (link) του γραφήματος γειτνίασης.
- Η μέθοδος αυτή ονομάζεται και **μέθοδος πλήρους συνδεσιμότητας (complete linkage)**.

# Απόσταση Συστάδων – Μέγιστη Απόσταση

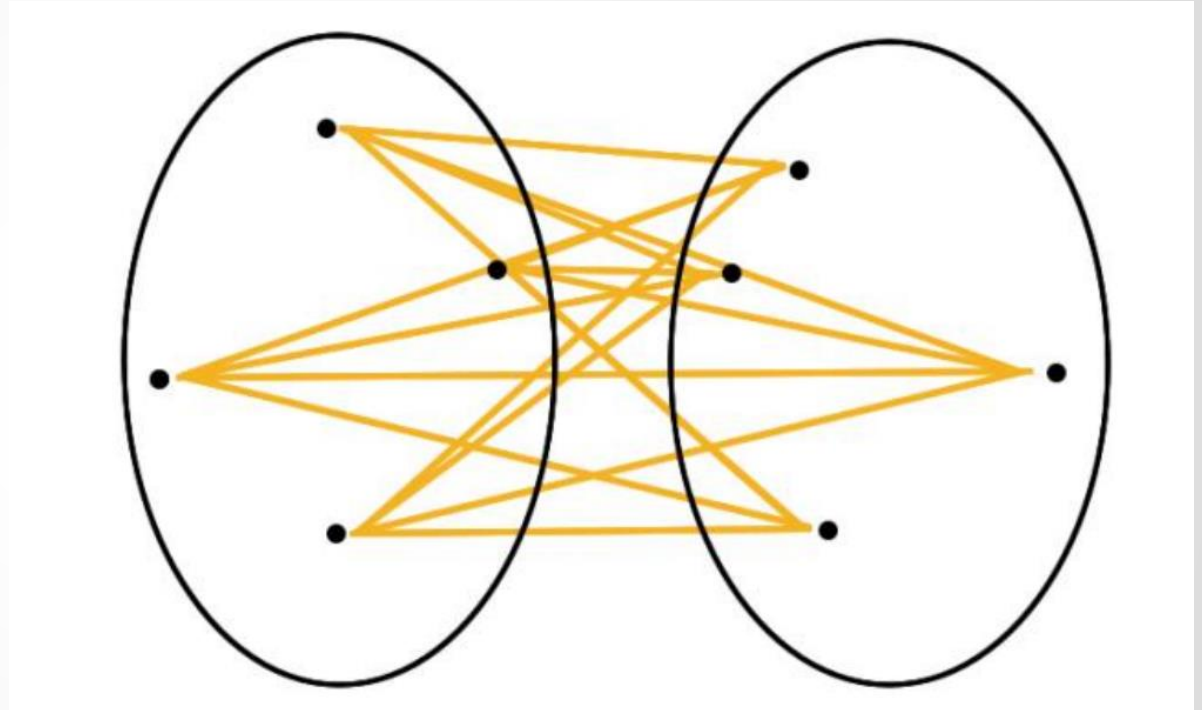
- Το βασικό πλεονέκτημα αυτού του τρόπου σύνδεσης είναι η μικρή ευαισθησία στον θόρυβο και στις ακραίες τιμές (outliers).
- Τα μειονεκτήματα που έχει είναι ότι τείνει να διασπά μεγάλες συστάδες και ότι οδηγεί, συνήθως, σε κυκλικά σχήματα.

# Απόσταση Συστάδων – Μέσος Όρος Συστάδων

- **Μέσος όρος Συστάδων:** Είναι η μέση τιμή των αποστάσεων όλων των διαφορετικών ζευγαριών μεταξύ των σημείων των δύο συστάδων.

# Απόσταση Συστάδων

Μέσος όρος συστάδων, ως μέτρο  
απόστασης δύο συστάδων



# Απόσταση Συστάδων – Μέσος Όρος Συστάδων

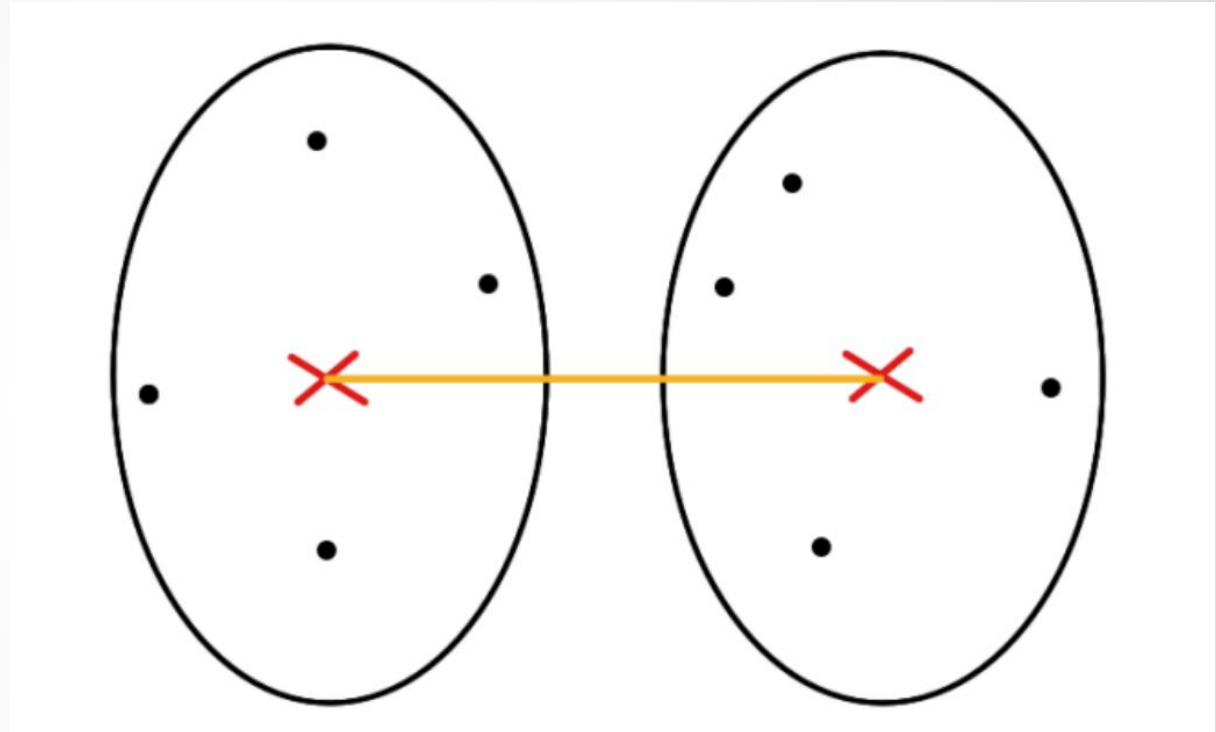
- Ο μέσος όρος συστάδων βρίσκεται κάπου ανάμεσα στην ελάχιστη και τη μέγιστη απόσταση.
- Έχει μικρότερη ευαισθησία σε θόρυβο και σε ακραίες τιμές (outliers), αλλά ευνοεί τις συστάδες με κυκλικό σχήμα.

# Απόσταση Συστάδων – Απόσταση Κεντρικών Σημείων

- **Απόσταση Κεντρικών Σημείων:** Είναι η απόσταση μεταξύ των κέντρων των συστάδων

# Απόσταση Συστάδων

Απόσταση κεντρικών σημείων, ως  
μέτρο απόστασης δύο συστάδων



# Απόσταση Συστάδων – Απόσταση Κεντρικών Σημείων

- Αυτού του είδους η απόσταση δεν έχει μονότονη αύξηση. Έτσι, δύο συστάδες που συγχωνεύονται μπορεί να έχουν μικρότερη απόσταση από συστάδες, οι οποίες έχουν συγχωνευτεί σε προηγούμενα βήματα.

# Απόσταση Συστάδων – Μέθοδος του Ward

Με την μέθοδο του Ward η απόσταση μεταξύ δύο συστάδων,  $C_i, C_j$  είναι ίση με το πόσο θα αυξηθεί το άθροισμα των τετραγώνων της απόστασης των στοιχείων της κάθε συστάδας από το αντίστοιχο κεντροειδές (της κάθε συστάδας) μετά τη συγχώνευση τους,  $C_{ij}$ .

# Είδη Ιεραρχικής Συσταδοποίησης

- Οι αλγόριθμοι ιεραρχικής συσταδοποίησης μπορεί να είναι είτε
  - **Συσσωρευτικοί Αλγόριθμοι (Agglomerative Algorithms)**
  - είτε
  - **Διαιρετικοί Αλγόριθμοι (Divisive Algorithms)**

# Συσσωρευτικοί Αλγόριθμοι

- Οι συσσωρευτικοί αλγόριθμοι ξεκινάνε με κάθε ένα από τα  $n$  δείγματα να ανήκει σε μια ξεχωριστή συστάδα, δηλαδή ξεκινάνε με  $n$  συστάδες.
- Σε κάθε βήμα, συγχωνεύονται οι δύο πιο κοντινές συστάδες, δηλαδή το πλήθος των συστάδων μειώνεται κατά ένα.

# Συσσωρευτικοί Αλγόριθμοι

- Αυτή η διαδικασία επαναλαμβάνεται, μέχρις ότου ο αλγόριθμος καταλήξει σε μια μοναδική συστάδα, η οποία θα εμπεριέχει όλα τα  $n$  δείγματα.
- Η όλη διαδικασία του αλγορίθμου μπορεί να αναπαρασταθεί με δενδρόγραμμα ανομοιοότητας. Το δενδρόγραμμα περιέχει  $n - 1$  επίπεδα και το κάθε επίπεδο αντιστοιχεί σε ένα βήμα του αλγορίθμου.

# Διαιρετικοί Αλγόριθμοι

- Οι διαιρετικοί αλγόριθμοι ξεκινάνε με όλα τα δείγματα να ανήκουν σε μια ενιαία συστάδα.
- Σε κάθε βήμα, μια ομάδα διασπάται σε δύο. Αυτό γίνεται επαναληπτικά, μέχρι να καταλήξουμε σε  $n$  ομάδες.

# Διαιρετικοί Αλγόριθμοι

- Η πολυπλοκότητα των διαιρετικών αλγορίθμων είναι μεγαλύτερη από αυτή των συσσωρευτικών, αφού η διάσπαση μιας ομάδας σε δύο μπορεί να γίνει κατά  $(2^{n-1} - 1)$  τρόπους.
- Η επιλογή της βέλτιστης διάσπασης πρακτικά είναι αδύνατη ακόμα και για μικρό  $n$ .

# Διαιρετικοί Αλγόριθμοι

- Στην πράξη η διάσπαση γίνεται, αλλά όχι κατά τον βέλτιστο τρόπο.
- Η όλη διαδικασία του διαιρετικού αλγορίθμου μπορεί να αναπαρασταθεί, όπως και στους συσσωρευτικούς, με δένδρoγραμμα.

# Παράδειγμα 1

- Το παρακάτω με λίγα δεδομένα και μεταβλητές παράδειγμα βοηθάει στην κατανόηση του συσσωρευτικού αλγόριθμου της ιεραρχικής συσταδοποίησης.



# Παράδειγμα 1

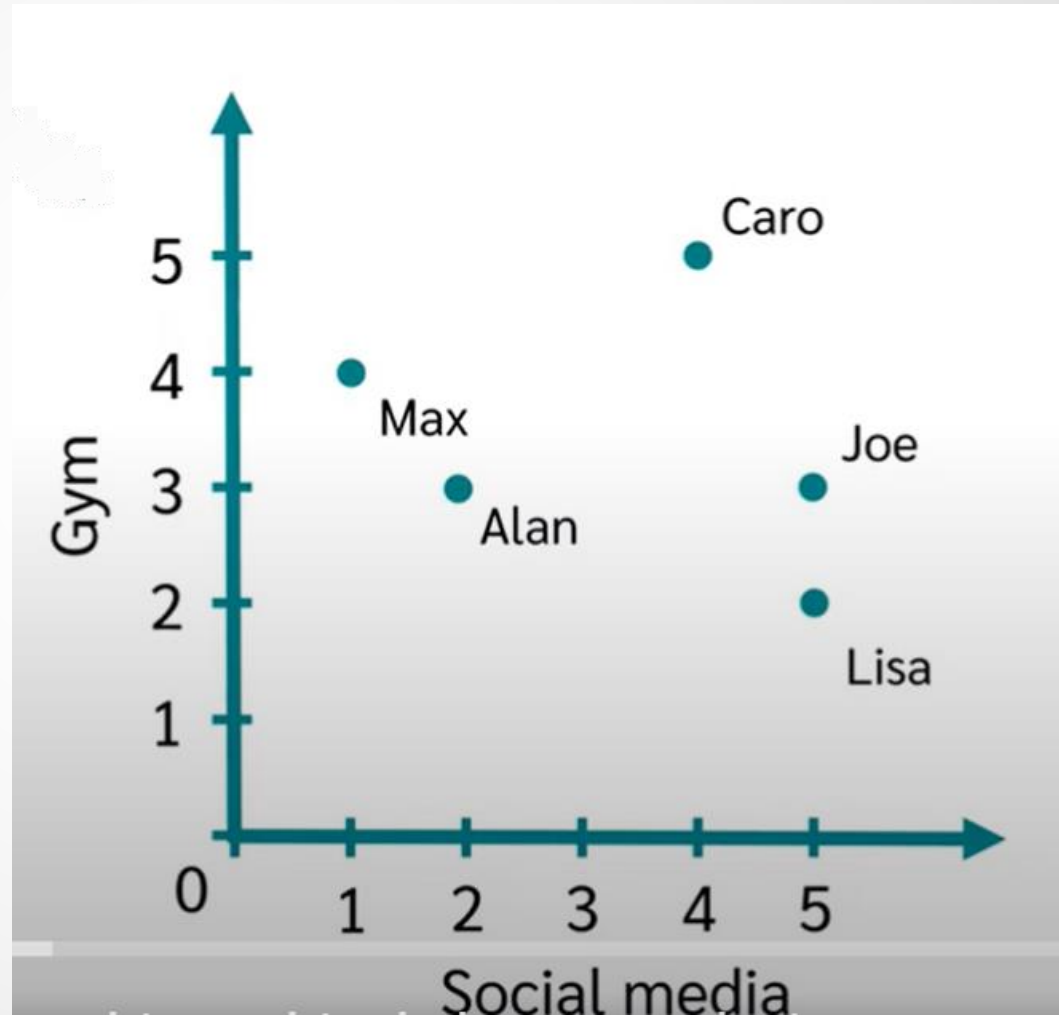
- Ρωτήθηκαν 5 άτομα ως προς τον χρόνο σε ώρες που αφιερώνουν την εβδομάδα σε γυμναστική και μέσα δικτύωσης.
- Οι απαντήσεις τους φαίνονται στον παρακάτω πίνακα

# Παράδειγμα 1

| Name | Social Media | Gym |
|------|--------------|-----|
| Alan | 2            | 3   |
| Lisa | 5            | 2   |
| Joe  | 5            | 3   |
| Max  | 1            | 4   |
| Caro | 4            | 5   |

# Παράδειγμα 1

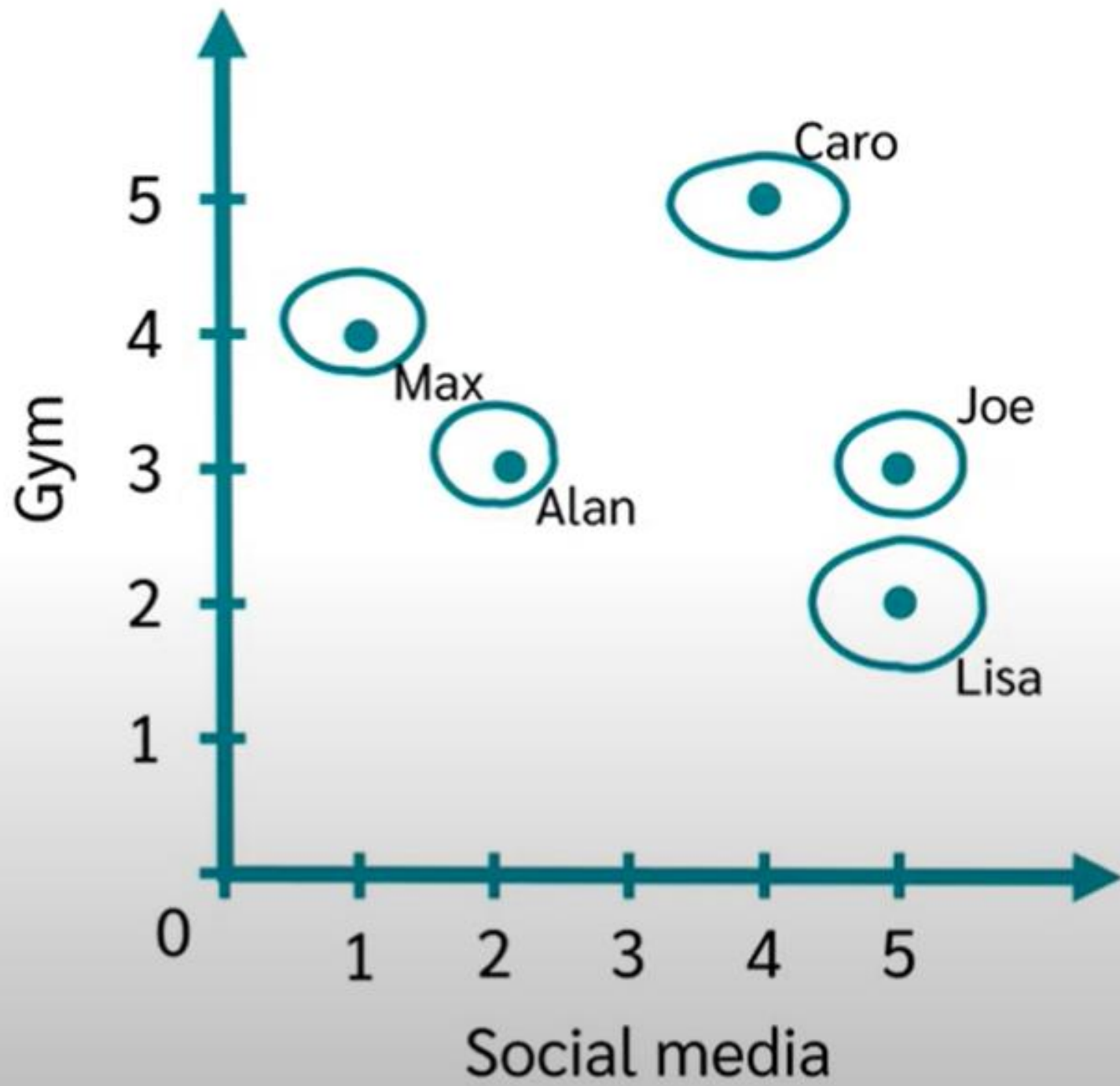
- Δημιουργούμε αρχικά διάγραμμα διασποράς με τα δεδομένα μας.



# Παράδειγμα 1

- Θεωρούμε στο πρώτο βήμα ότι κάθε στοιχείο του δείγματος ανήκει σε μία μοναδική συστάδα.
- **Επομένως έχουμε τόσα clusters όσα και το μέγεθος του δείγματος,** (στο συγκεκριμένο παράδειγμα 5 clusters).

# Παράδειγμα 1

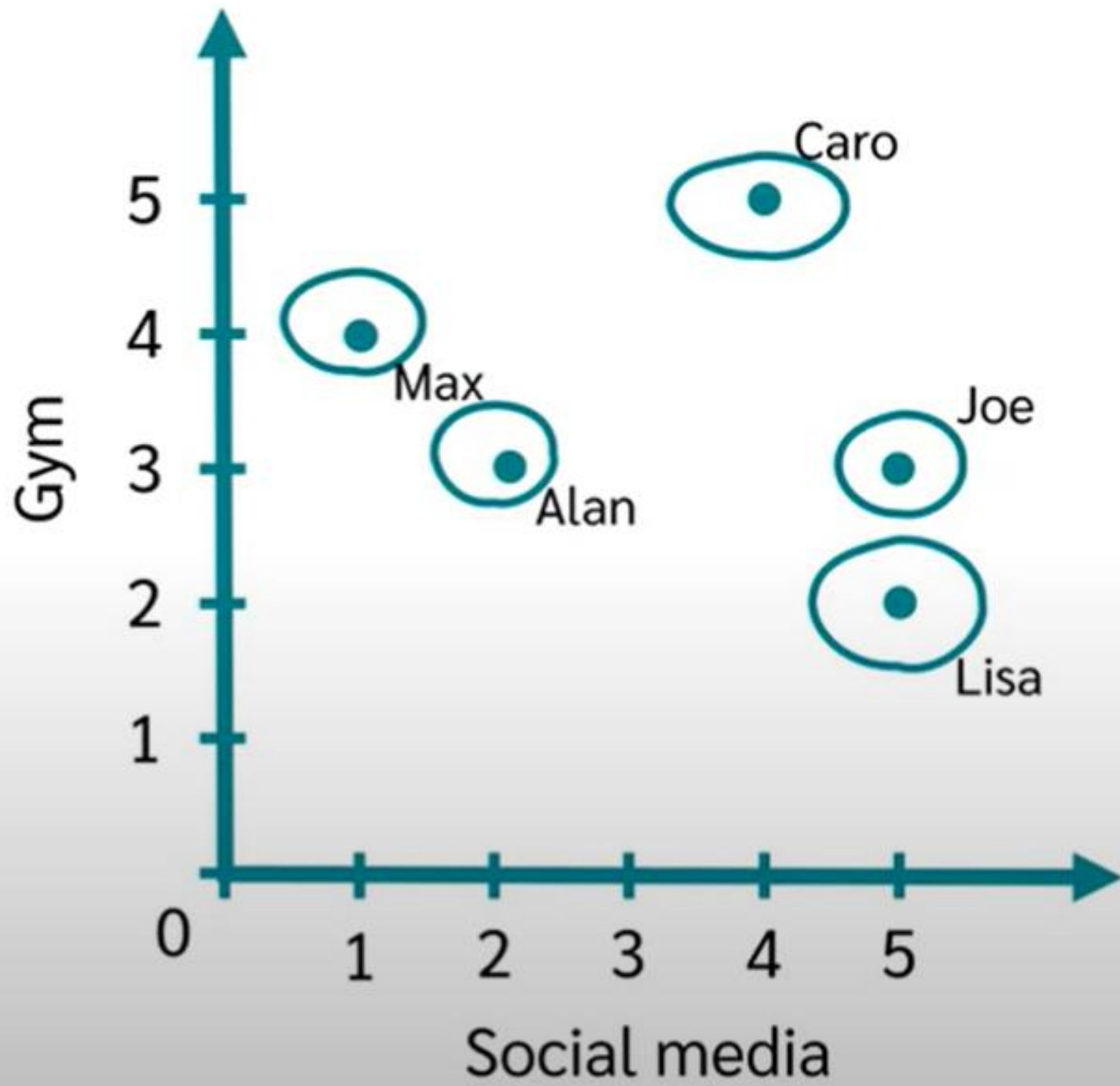


# Παράδειγμα 1

- Ο σκοπός είναι να συγχωνεύσουμε όλο και περισσότερες συστάδες μέχρι το σημείο που όλα τα στοιχεία του δείγματος θα ανήκουν σε ένα μοναδικό cluster.



# Παράδειγμα 1

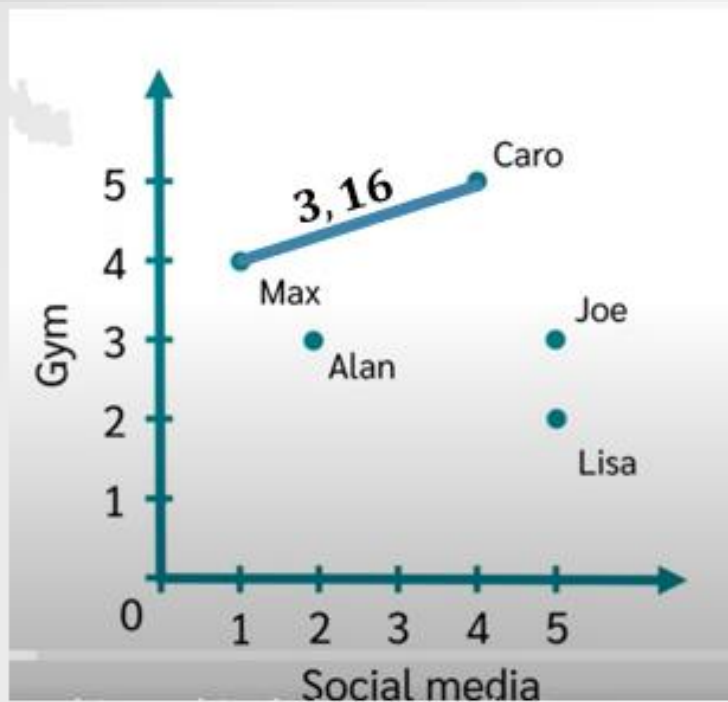


# Παράδειγμα 1

- Σε κάθε βήμα τα δύο clusters με την μικρότερη μεταξύ τους απόσταση συγχωνεύονται σε ένα.
- Η απόσταση μετριέται με την Ευκλείδεια απόσταση.
- Η απόσταση μεταξύ συστάδων μπορεί να μετρηθεί με έναν από τους τρόπους που έχουμε ήδη αναφέρει. Στο παράδειγμα μας θα χρησιμοποιήσουμε την μέθοδο της ελάχιστης απόστασης.

# Παράδειγμα 1

- Στο πρώτο βήμα έχουμε ότι κάθε στοιχεί ανήκει και σε διαφορετική συστάδα. Επομένως οι αποστάσεις μεταξύ συστάδων είναι ίσες με τις αποστάσεις μεταξύ των στοιχείων αντίστοιχα.
- Δημιουργούμε των πίνακα αποστάσεων μεταξύ των στοιχείων του δείγματος.



|      | Alan | Lisa | Joe  | Max  | Caro |
|------|------|------|------|------|------|
| Alan | 0    |      |      |      |      |
| Lisa | 3,16 | 0    |      |      |      |
| Joe  | 3,00 | 1,00 | 0    |      |      |
| Max  | 1,41 | 4,47 | 4,12 | 0    |      |
| Caro | 2,83 | 3,16 | 2,24 | 3,16 | 0    |

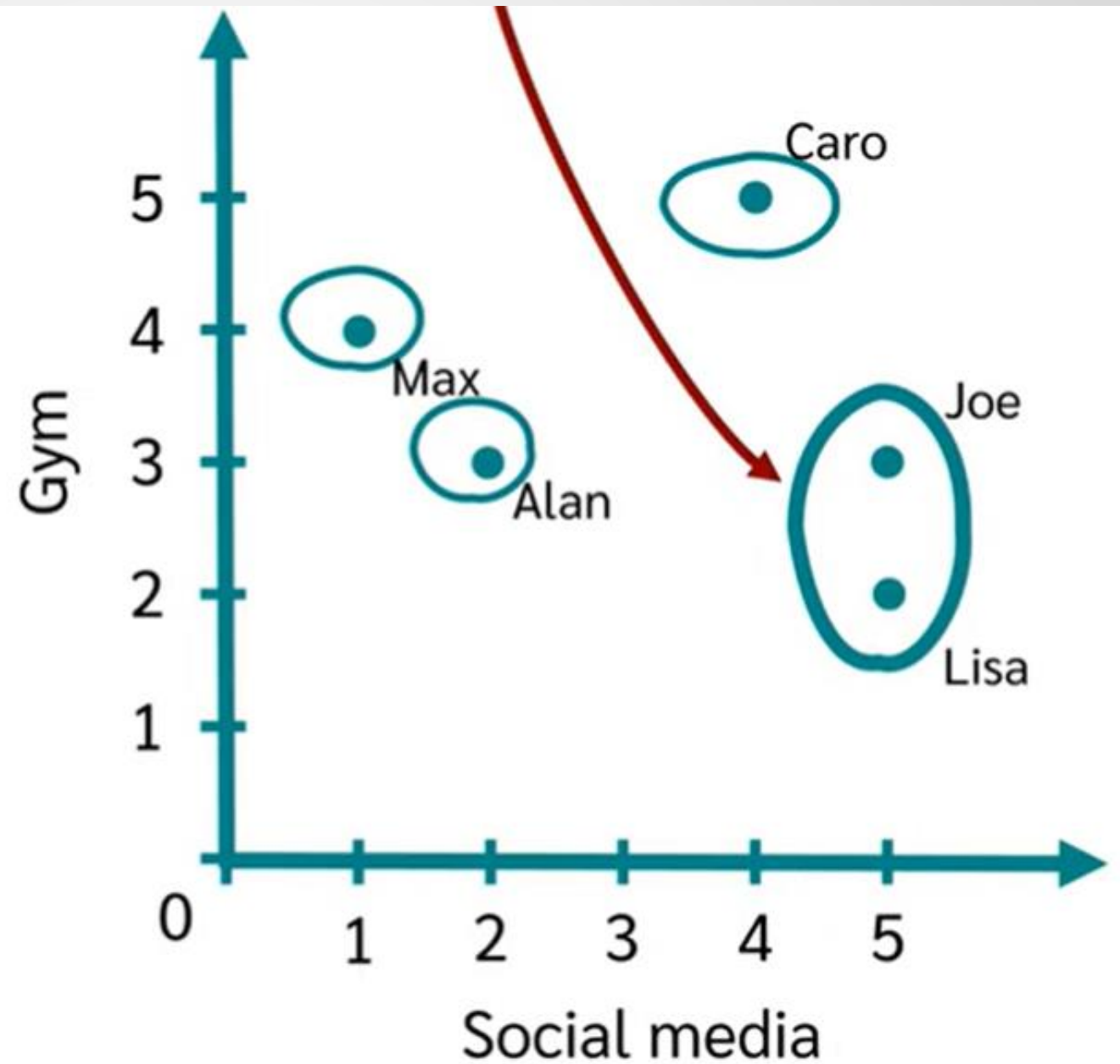
# Παράδειγμα 1

# Παράδειγμα 1

- Όπως φαίνεται από τον πίνακα αποστάσεων, η μικρότερη απόσταση, είναι μεταξύ των συστάδων “Joe” και “Lisa”. Επομένως αυτές οι δύο συστάδες θα συγχωνευτούν στο δεύτερο βήμα του αλγόριθμου.

|      | Alan | Lisa | Joe  | Max  | Caro |
|------|------|------|------|------|------|
| Alan | 0    |      |      |      |      |
| Lisa | 3,16 | 0    |      |      |      |
| Joe  | 3,00 | 1,00 | 0    |      |      |
| Max  | 1,41 | 4,47 | 4,12 | 0    |      |
| Caro | 2,83 | 3,16 | 2,24 | 3,16 | 0    |

# Παράδειγμα 1



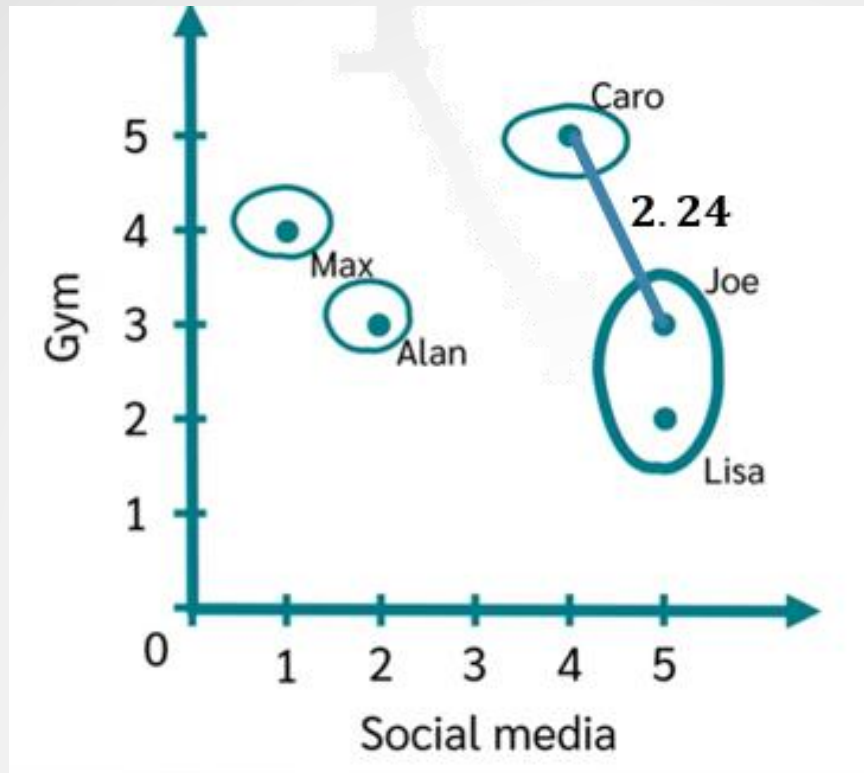
# Παράδειγμα 1

- Στη συνέχεια πρέπει να κατασκευάσουμε τον πίνακα αποστάσεων των νέων clusters.
- Θα χρησιμοποιήσουμε την μέθοδο της ελάχιστης απόστασης.
- Βλέπουμε παρακάτω τον πίνακα των αποστάσεων των νέων clusters.



|           | Alan | Lisa, Joe | Max  | Caro |
|-----------|------|-----------|------|------|
| Alan      | 0    |           |      |      |
| Lisa, Joe | 3,00 | 0         |      |      |
| Max       | 1,41 | 4,12      | 0    |      |
| Caro      | 2,83 | 2,24      | 3,16 | 0    |

## Παράδειγμα 1



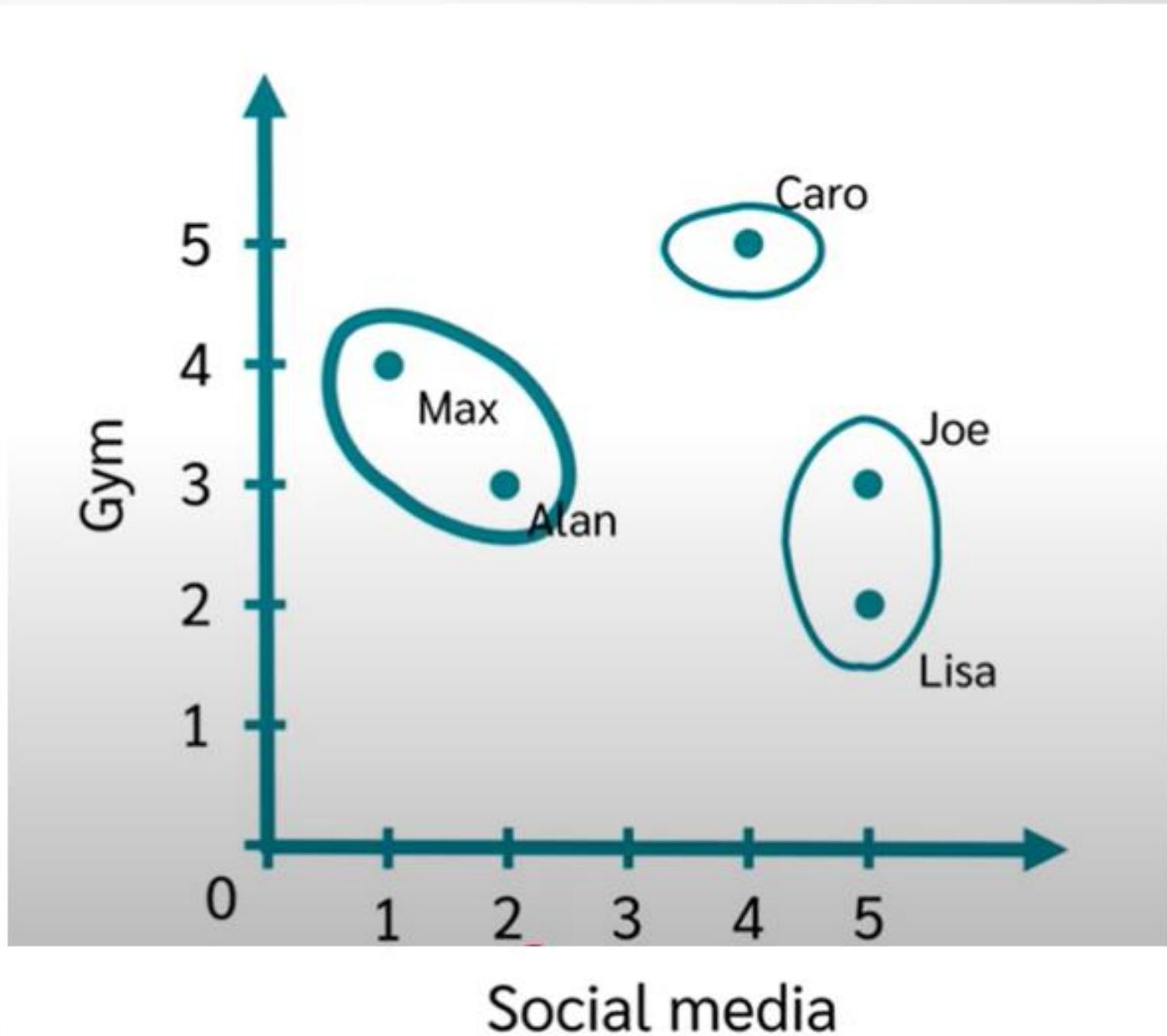
# Παράδειγμα 1

# Παράδειγμα 1

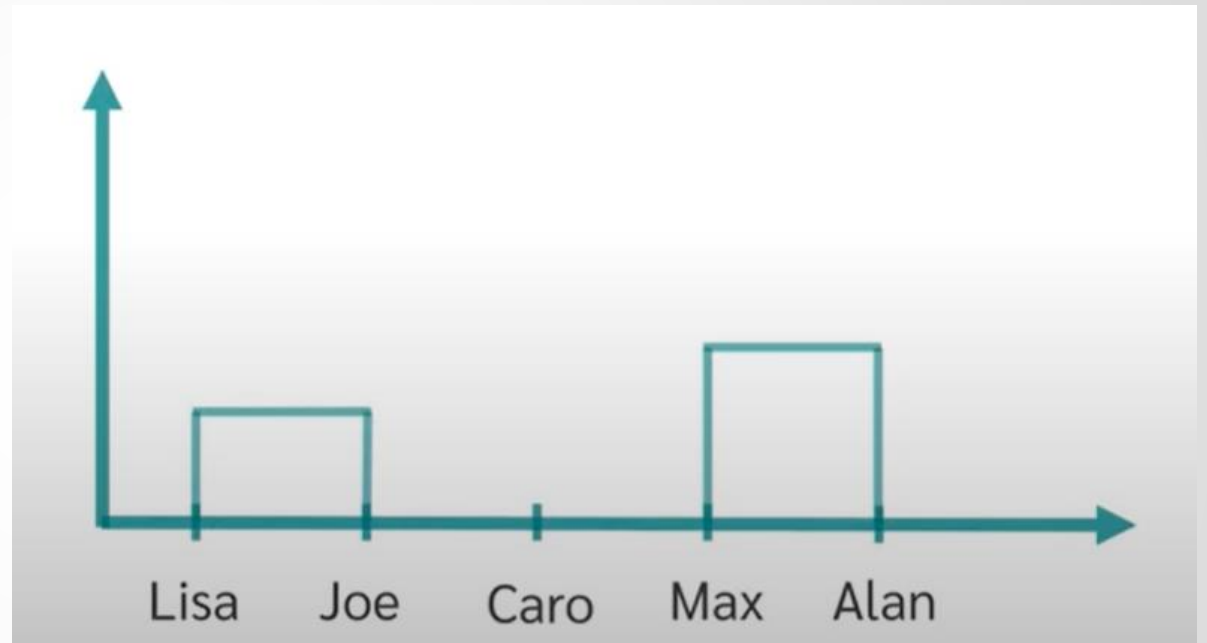
- Όπως φαίνεται από τον πίνακα αποστάσεων, η μικρότερη απόσταση είναι μεταξύ των συστάδων “Max” και “Alan”.
- Επομένως αυτές οι δύο συστάδες θα συγχωνευτούν στο επόμενο βήμα του αλγόριθμου.

|           | Alan | Lisa, Joe | Max  | Caro |
|-----------|------|-----------|------|------|
| Alan      | 0    |           |      |      |
| Lisa, Joe | 3,00 | 0         |      |      |
| Max       | 1,41 | 4,12      | 0    |      |
| Caro      | 2,83 | 2,24      | 3,16 | 0    |

# Παράδειγμα 1



# Παράδειγμα 1



# Παράδειγμα 1

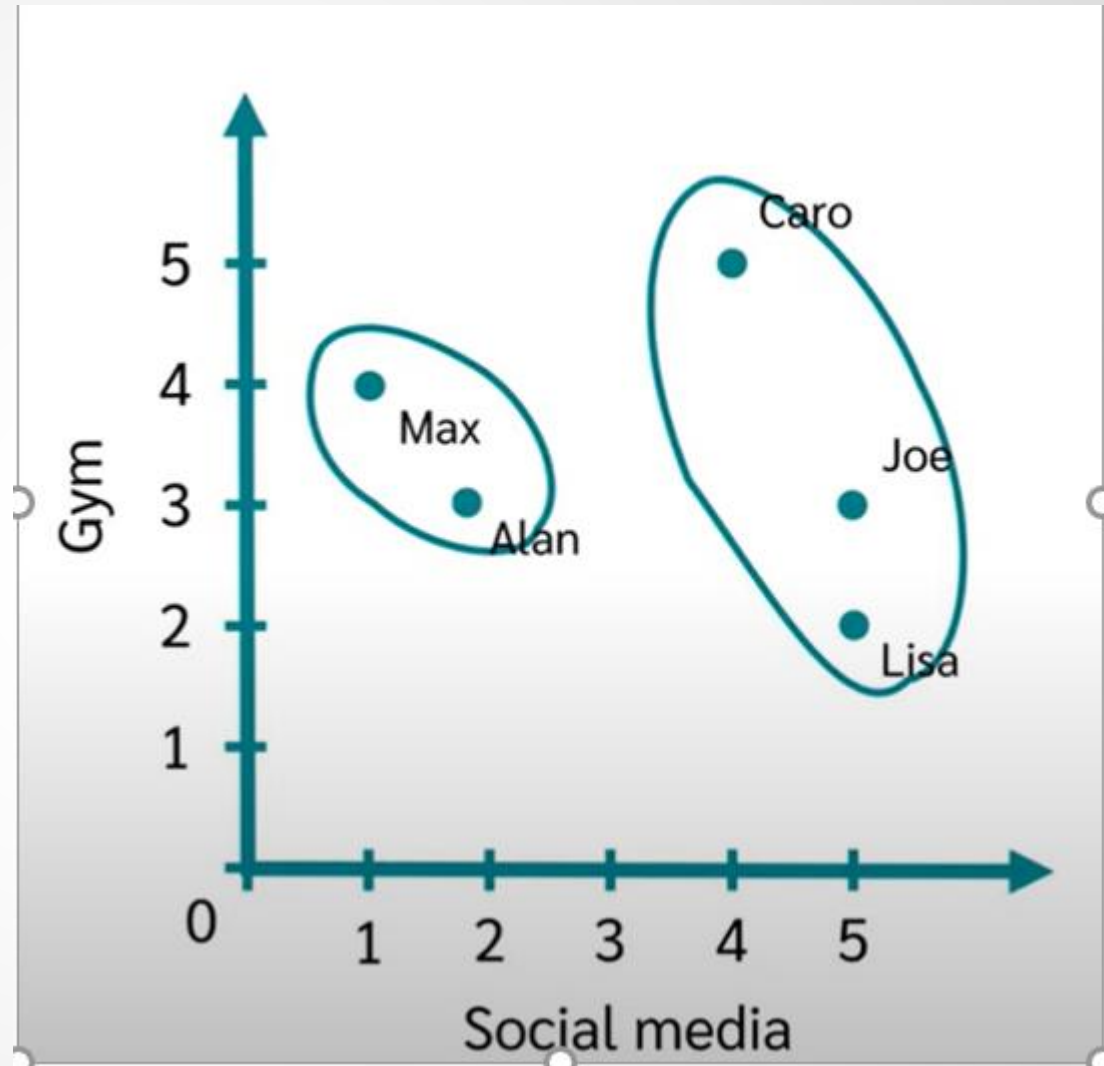
- Στη συνέχεια κατασκευάζουμε εκ νέου τον πίνακα αποστάσεων των νέων clusters και συγχωνεύουμε τις συστάδες με την μικρότερη απόσταση.



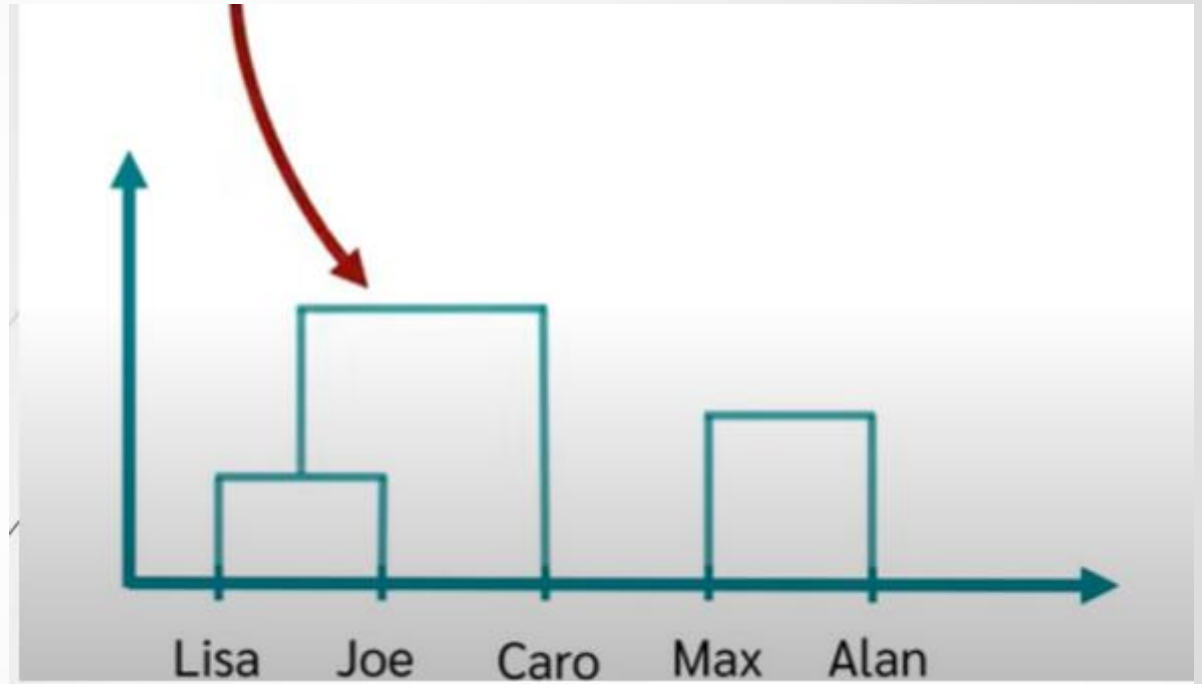
# Παράδειγμα 1

|           | Lisa, Joe | Max, Alan | Caro |
|-----------|-----------|-----------|------|
| Lisa, Joe | 0         |           |      |
| Max, Alan | 3,00      | 0         |      |
| Caro      | 2,24      | 2,83      | 0    |

# Παράδειγμα 1



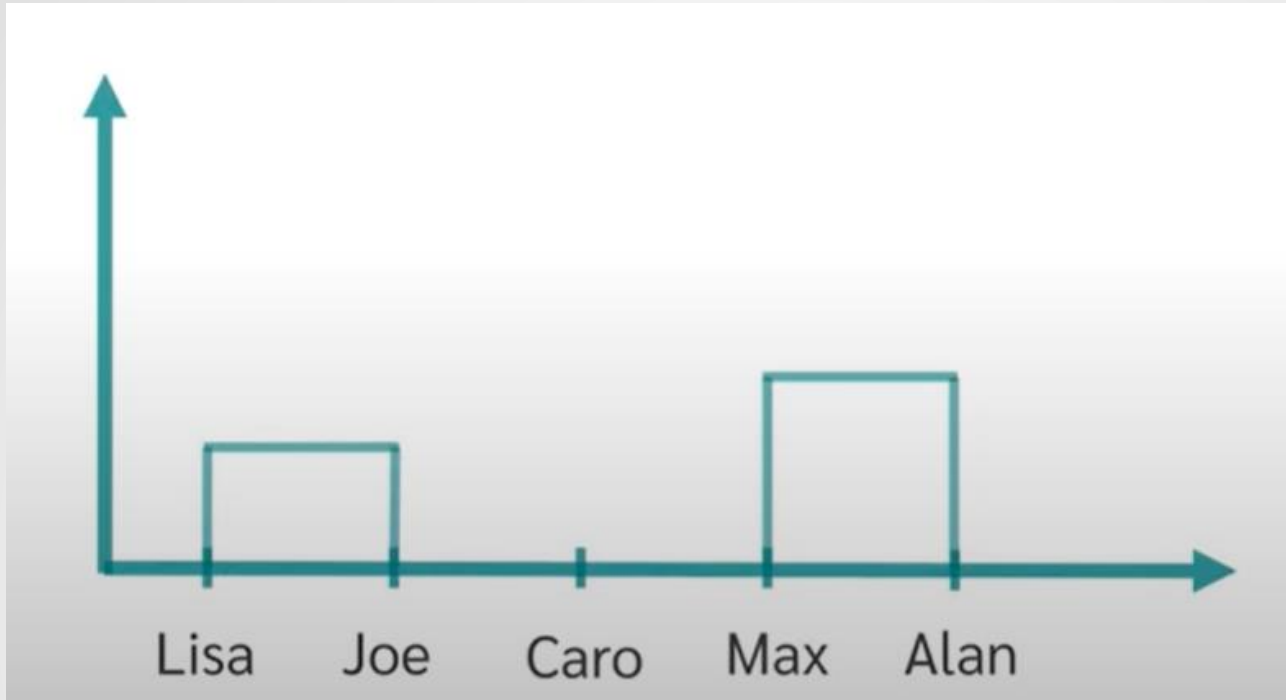
# Παράδειγμα 1



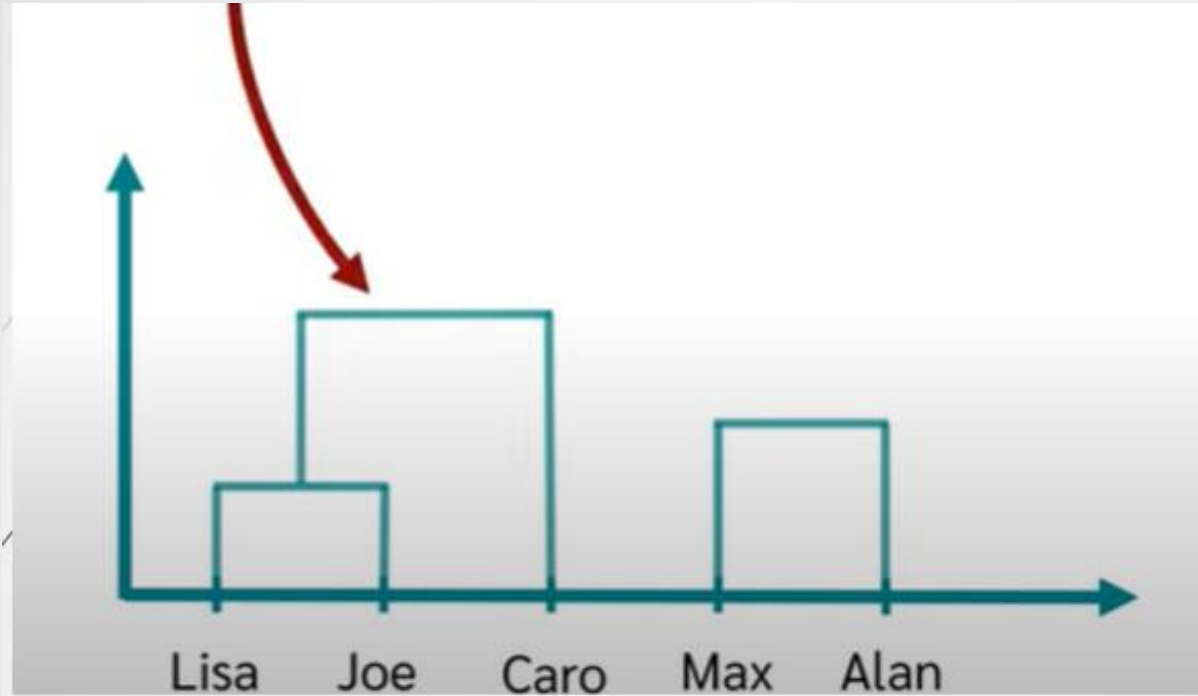
# Παράδειγμα 1

- Βλέπουμε στα παρακάτω σχήματα και την κατασκευή του δεντροδιαγράμματος σε κάθε βήμα.

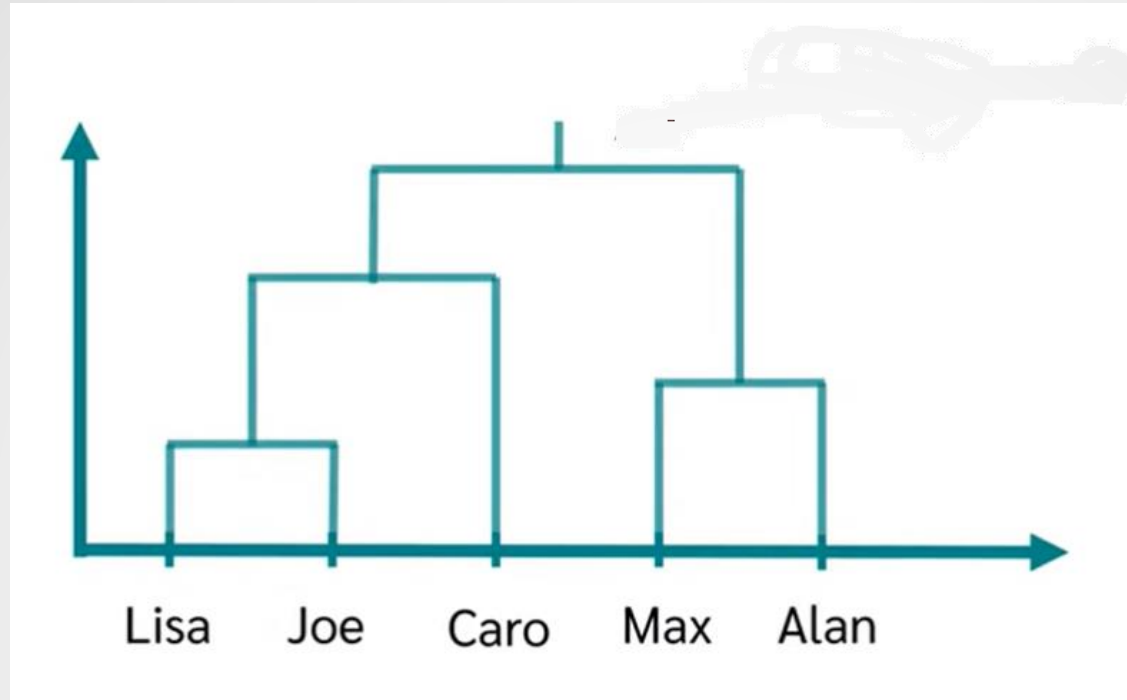




## Παράδειγμα 1



## Παράδειγμα 1



## Παράδειγμα 1

# Βήματα Αλγόριθμου Ierarchical Clustering – Σύνοψη

Στη γενική περίπτωση τα βήματα του συσσωρευτικού αλγόριθμου ιεραρχικής συσταδοποίησης μπορούν να συνοψιστούν ως εξής:

- **Βήμα 1:** Κατασκευάζουμε τον πίνακα αποστάσεων των σημείων του δείγματος, χρησιμοποιώντας μία κατάλληλη μετρική.
- **Βήμα 2:** Κάθε σημείο του δείγματος ανήκει σε ένα μοναδικό cluster

# Βήματα Αλγόριθμου Ierarchical Clustering – Σύνοψη

- **Βήμα 3:** Συγχωνεύουμε τα δύο clusters με την μικρότερη απόσταση.
- **Βήμα 4:** Κατασκευάζουμε τον νέο πίνακα αποστάσεων μεταξύ των νέων clusters.

# Βήματα Αλγόριθμου Ierarchical Clustering – Σύνοψη

- **Βήμα 5:** Επαναλαμβάνουμε τα βήματα 3 και 4 μέχρι το σημείο που θα έχουμε ένα και μοναδικό cluster στο οποίο θα ανήκουν όλα τα σημεία.

## Παράδειγμα 2

Στον παρακάτω πίνακα έχουμε δεδομένα από 36 πελάτες ενός φυσικού καταστήματος.

Για κάθε έναν από τους πελάτες έχουμε δεδομένα για το φύλο του (1: Άντρας, 0: Γυναίκα), το έσοδα σε € που έχει το κατάστημα από τον πελάτη αυτόν (Total Purchase) καθώς και το ετήσιο εισόδημα του πελάτη (Yearly Revenue).

## Παράδειγμα 2

|    | A    | B              | C              | D      | E |
|----|------|----------------|----------------|--------|---|
| 1  |      |                |                |        |   |
| 2  | Name | Total Purchase | Yearly Revenue | Gender |   |
| 3  | A    | 5.000          | 100.000        | 1      |   |
| 4  | AA   | 1.700          | 25.000         | 2      |   |
| 5  | B    | 6.000          | 85.000         | 2      |   |
| 6  | BB   | 2.000          | 20.000         | 1      |   |
| 7  | C    | 7.000          | 105.000        | 2      |   |
| 8  | CC   | 3.000          | 30.000         | 1      |   |
| 9  | D    | 8.000          | 75.000         | 1      |   |
| 10 | DD   | 4.000          | 20.000         | 2      |   |
| 11 | E    | 9.000          | 110.000        | 1      |   |
| 12 | EE   | 5.000          | 43.000         | 2      |   |
| 13 | F    | 7.500          | 100.000        | 1      |   |
| 14 | FF   | 6.000          | 25.000         | 2      |   |
| 15 | G    | 6.500          | 105.000        | 2      |   |
| 16 | GG   | 7.000          | 30.000         | 2      |   |
| 17 | H    | 6.000          | 95.000         | 1      |   |
| 18 | HH   | 8.000          | 20.000         | 2      |   |
| 19 | I    | 6.500          | 110.000        | 1      |   |
| 20 | II   | 9.000          | 60.000         | 2      |   |
| 21 | J    | 7.000          | 100.000        | 2      |   |
| 22 | JJ   | 10.000         | 25.000         | 2      |   |
| 23 | K    | 7.500          | 105.000        | 2      |   |
| 24 | L    | 18.000         | 95.000         | 1      |   |
| 25 | M    | 16.000         | 110.000        | 2      |   |
| 26 | N    | 12.000         | 100.000        | 1      |   |
| 27 | O    | 11.000         | 105.000        | 1      |   |
| 28 | P    | 20.000         | 95.000         | 1      |   |
| 29 | Q    | 21.000         | 110.000        | 1      |   |
| 30 | R    | 22.000         | 100.000        | 1      |   |
| 31 | S    | 23.000         | 105.000        | 1      |   |
| 32 | T    | 24.000         | 95.000         | 2      |   |
| 33 | U    | 25.000         | 110.000        | 2      |   |
| 34 | V    | 26.000         | 100.000        | 2      |   |
| 35 | W    | 27.000         | 105.000        | 2      |   |
| 36 | X    | 28.000         | 95.000         | 2      |   |
| 37 | Y    | 29.000         | 110.000        | 1      |   |
| 38 | Z    | 30.000         | 100.000        | 2      |   |
| 39 |      |                |                |        |   |
| 40 |      |                |                |        |   |

## Παράδειγμα 2

Προκειμένου ο ιδιοκτήτης του φυσικού καταστήματος να οργανώσει αποτελεσματικότερα τις επιχειρηματικές του επενδύσεις θέλει να ομαδοποιήσει τους παραπάνω 36 πελάτες σε συστάδες ως προς τα έσοδα σε € που έχει το κατάστημα από τον πελάτη αυτόν (Total Purchase) και το ετήσιο εισόδημα του πελάτη (Yearly Revenue).



## Παράδειγμα 2 – Υλοποίηση Hierarchical Clustering στο SPSS

- Θέλουμε να ομαδοποιήσουμε τους πελάτες του φυσικού καταστήματος με βάση τις τιμές των μεταβλητών **«Total Purchase»** και **«Yearly Revenue»**.

## Παράδειγμα 2 – Υλοποίηση Hierarchical Clustering στο SPSS

- Στην ιεραρχική συσταδοποίηση δεν χρειάζεται να ορίσουμε έναν συγκεκριμένο αριθμό συστάδων.
- Οποιοσδήποτε αριθμός μπορεί να επιτευχθεί, κόβοντας το δένδροδιάγραμμα σε ένα κατάλληλο επίπεδο.

## Παράδειγμα 2 – Υλοποίηση Hierarchical Clustering στο SPSS

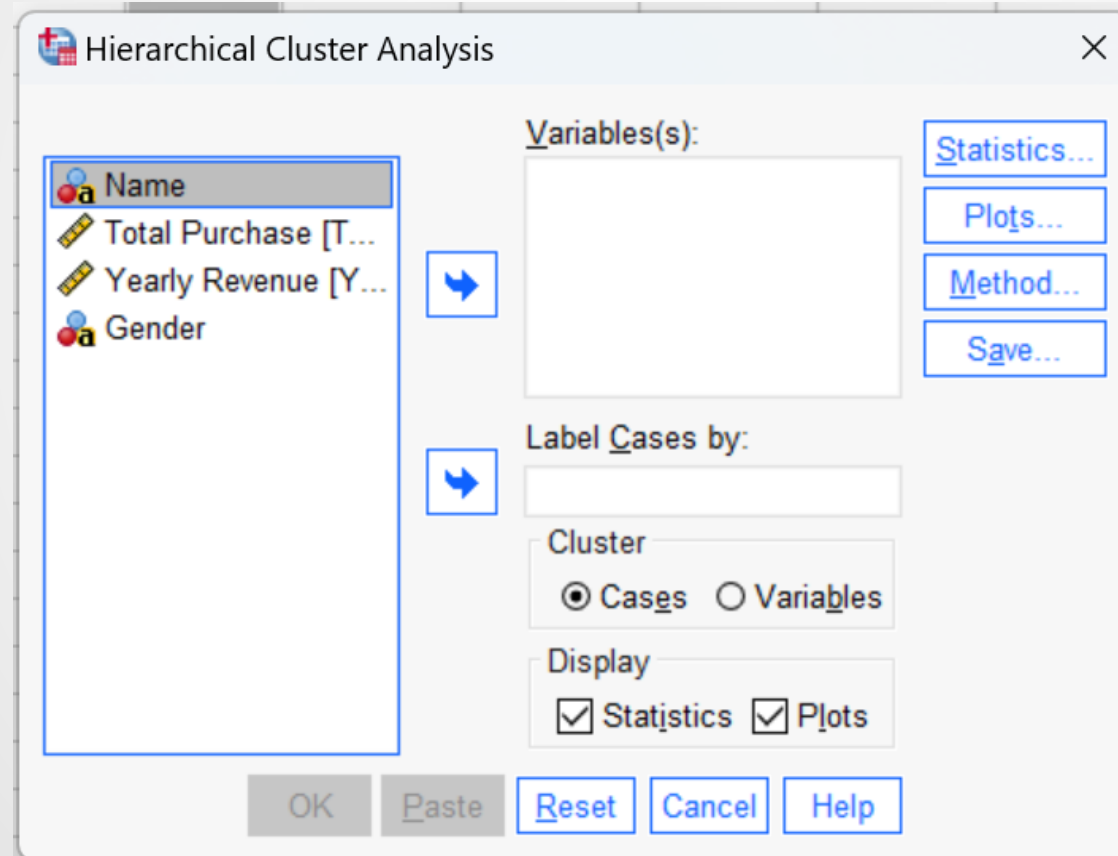
- Η συσταδική ανάλυση πραγματοποιείται ακολουθώντας την διαδρομή

*Analyse → Classify → Hierarchical Clusters*

Στη συνέχεια το SPSS ανοίγει το παρακάτω παράθυρο διαλόγου.



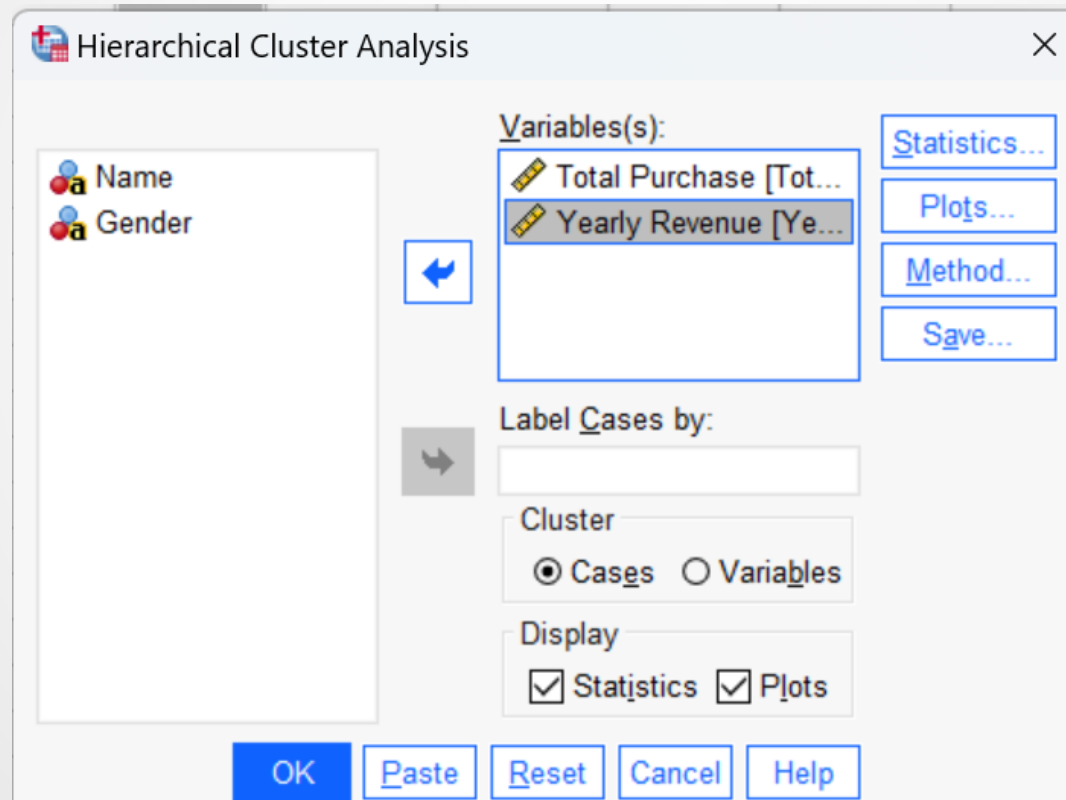
# Υλοποίηση Hierarchical Clustering στο SPSS



# Υλοποίηση Hierarchical Clustering στο SPSS

- Στο παράθυρο διαλόγου πρέπει να δηλώσουμε:
  - Τις μεταβλητές βάση των οποίων θα πραγματοποιηθεί η συσταδική ανάλυση **(τις τοποθετούμε στο πλαίσιο Variables)**.

# Υλοποίηση Hierarchical Clustering στο SPSS



# Υλοποίηση Hierarchical Clustering στο SPSS

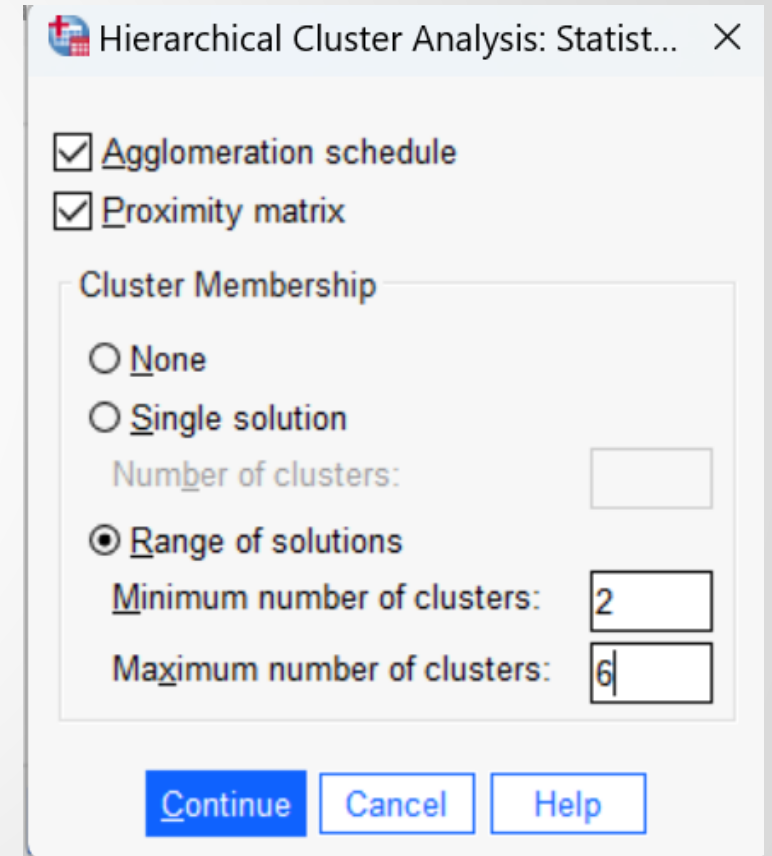
- Στο ίδιο παράθυρο διαλόγου και PIN πατήσουμε OK διαμορφώνουμε κατάλληλες ρυθμίσεις από τις τρεις επιλογές

**Statistics   Plots   Method   Save**

Κάθε μία από τις παραπάνω επιλογές ανοίγει και ένα νέο πλαίσιο διαλόγου όπως θα δούμε στα παρακάτω σχήματα.

# Υλοποίηση Hierarchical Clustering στο SPSS– Statistics

- Στο παράθυρο διαλόγου **Statistics** Επιλέγουμε την εμφάνιση **Proximity matrix**, και επιπλέον το εύρος των λύσεων (**Range of solutions**) που θέλουμε να μας εμφανίσει.



The screenshot shows the 'Hierarchical Cluster Analysis: Statistics' dialog box in SPSS. The 'Agglomeration schedule' and 'Proximity matrix' checkboxes are both checked. Under the 'Cluster Membership' section, the 'Range of solutions' radio button is selected. The 'Number of clusters' field is empty. The 'Minimum number of clusters' field contains the value '2', and the 'Maximum number of clusters' field contains the value '6'. At the bottom, there are three buttons: 'Continue', 'Cancel', and 'Help'.

Hierarchical Cluster Analysis: Statist... X

☒ Agglomeration schedule

☒ Proximity matrix

Cluster Membership

☐ None

☐ Single solution

Number of clusters:

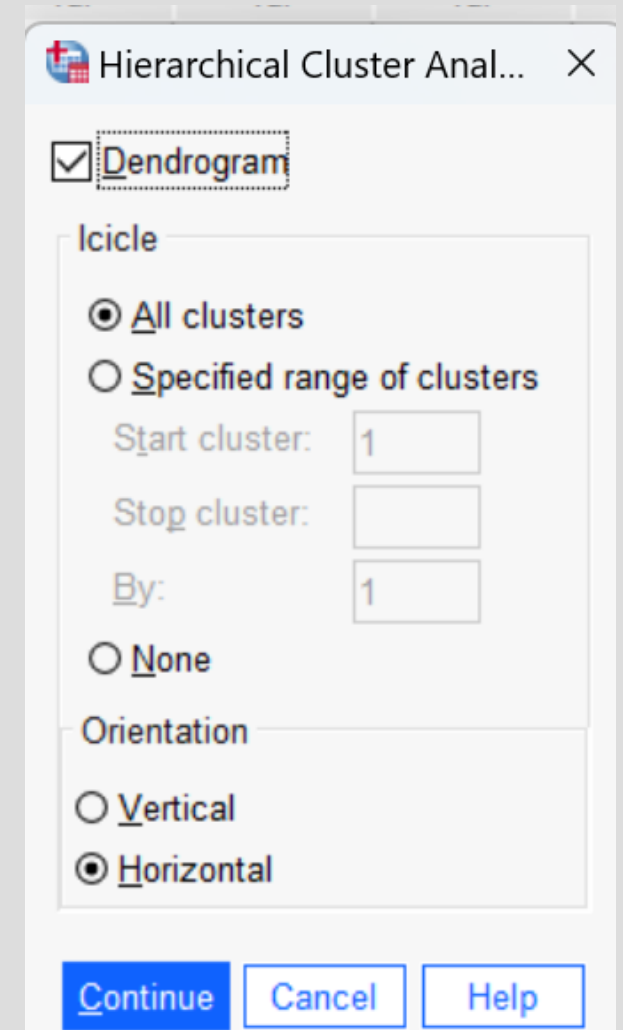
☒ Range of solutions

Minimum number of clusters:

Maximum number of clusters:

# Υλοποίηση Hierarchical Clustering στο SPSS– Plots

- Πατώντας το κουμπί επιλογών **Plots** επιλέγουμε την εμφάνιση δένδροδιαγράμματος καθώς και αν θέλουμε την δημιουργία όλων των δυνατών clusters ή συγκεκριμένο εύρος.



The screenshot shows the 'Hierarchical Cluster Anal...' dialog box in SPSS. The 'Plots' sub-dialog is active, showing the following options:

- ☒ **Dendrogram**
- Icicle**
  - ☒ **All clusters**
  - ☐ **Specified range of clusters**
    - Start cluster: 1
    - Stop cluster:
    - By: 1
  - ☐ **None**
- Orientation**
  - ☐ **Vertical**
  - ☒ **Horizontal**

At the bottom, there are three buttons: **Continue**, **Cancel**, and **Help**.

# Υλοποίηση Hierarchical Clustering στο SPSS– Method

- Τέλος, πατώντας το κουμπί επιλογών **Method** μπορούμε να επιλέξουμε:

The screenshot shows the 'Hierarchical Cluster Analysis: Method' dialog box in SPSS. The 'Cluster Method' is set to 'Nearest neighbor'. Under the 'Measure' section, 'Interval' is selected with 'Squared Euclidean distance', and 'Power' and 'Root' are both set to 2. 'Counts' is set to 'Chi-squared measure' and 'Binary' is set to 'Squared Euclidean distance'. Under 'Transform Values', 'Standardize' is set to 'None', and 'By variable' is selected. Under 'Transform Measure', 'Absolute values', 'Change sign', and 'Rescale to 0-1 range' are all unchecked. At the bottom are 'Continue', 'Cancel', and 'Help' buttons.

Hierarchical Cluster Analysis: Method

Cluster Method: Nearest neighbor

Measure

☒ Interval: Squared Euclidean distance

Power: 2 Root: 2

☐ Counts: Chi-squared measure

☐ Binary: Squared Euclidean distance

Present: 1 Absent: 0

Transform Values

Standardize: None

☒ By variable

☐ By case:

Transform Measure

☐ Absolute values

☐ Change sign

☐ Rescale to 0-1 range

Continue Cancel Help

# Υλοποίηση Hierarchical Clustering στο SPSS

## – Method

- Μέθοδος απόστασης μεταξύ Clusters.
- Μετρική απόστασης.

The screenshot shows the 'Hierarchical Cluster Analysis: Method' dialog box in SPSS. The 'Cluster Method' is set to 'Nearest neighbor'. Under the 'Measure' section, 'Interval' is selected with 'Squared Euclidean distance' as the measure. The 'Power' is set to 2 and the 'Root' is set to 2. 'Counts' and 'Binary' are not selected. Under 'Transform Values', 'Standardize' is set to 'None', and 'By variable' is selected. Under 'Transform Measure', 'Absolute values', 'Change sign', and 'Rescale to 0-1 range' are all unchecked. The 'Continue', 'Cancel', and 'Help' buttons are at the bottom.

Hierarchical Cluster Analysis: Method

Cluster Method: Nearest neighbor

Measure

☒ Interval: Squared Euclidean distance

Power: 2 Root: 2

☐ Counts: Chi-squared measure

☐ Binary: Squared Euclidean distance

Present: 1 Absent: 0

Transform Values

Standardize: None

☒ By variable

☐ By case:

Transform Measure

☐ Absolute values

☐ Change sign

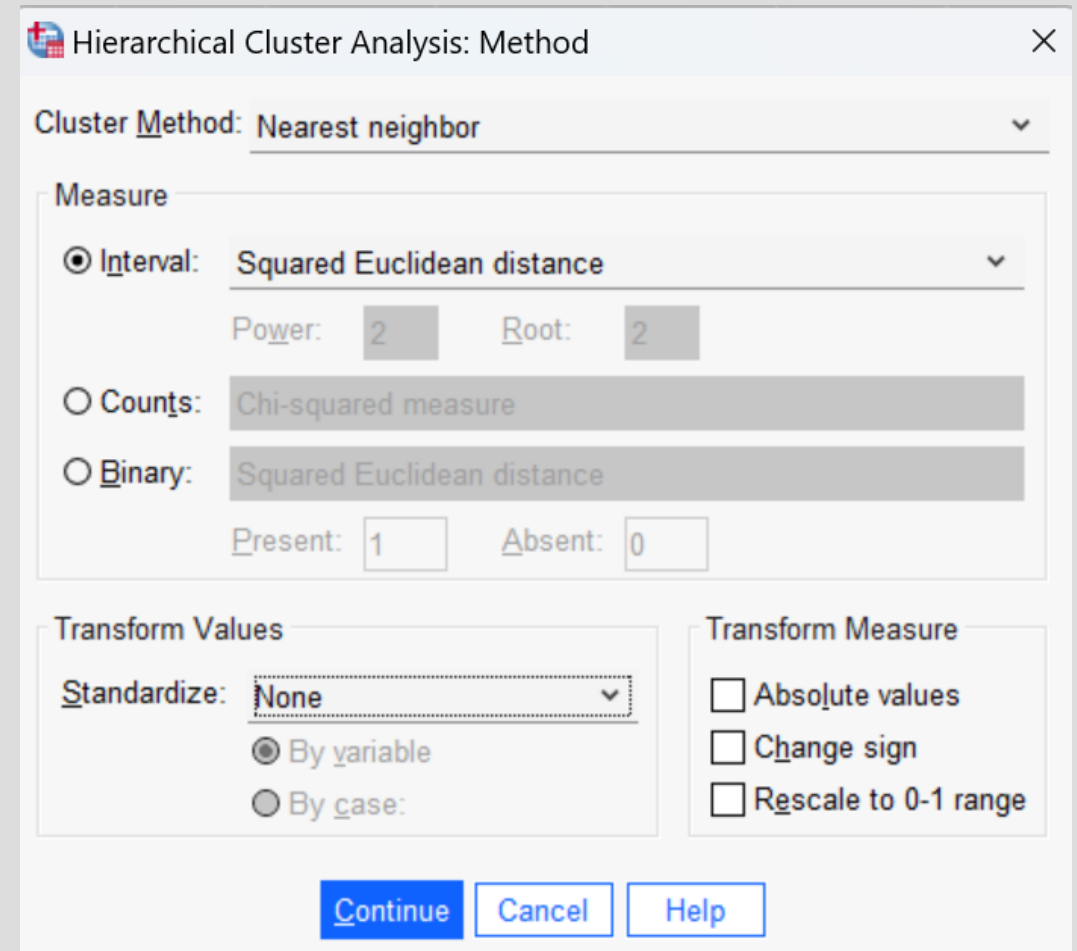
☐ Rescale to 0-1 range

Continue Cancel Help

# Υλοποίηση Hierarchical Clustering στο SPSS

## – Method

- Να μετατραπούν τα δεδομένα στην ίδια κλίμακα (standardize) ή όχι και με ποια μέθοδο.



The screenshot shows the 'Hierarchical Cluster Analysis: Method' dialog box in SPSS. The 'Cluster Method' is set to 'Nearest neighbor'. Under the 'Measure' section, 'Interval' is selected with 'Squared Euclidean distance', and 'Power' and 'Root' are both set to 2. 'Counts' is set to 'Chi-squared measure' and 'Binary' is set to 'Squared Euclidean distance'. Under 'Transform Values', 'Standardize' is set to 'None', and 'By variable' is selected. Under 'Transform Measure', 'Absolute values', 'Change sign', and 'Rescale to 0-1 range' are all unchecked. The 'Continue', 'Cancel', and 'Help' buttons are at the bottom.

Hierarchical Cluster Analysis: Method

Cluster Method: Nearest neighbor

Measure

☒ Interval: Squared Euclidean distance

Power: 2 Root: 2

☐ Counts: Chi-squared measure

☐ Binary: Squared Euclidean distance

Present: 1 Absent: 0

Transform Values

Standardize: None

☒ By variable

☐ By case:

Transform Measure

☐ Absolute values

☐ Change sign

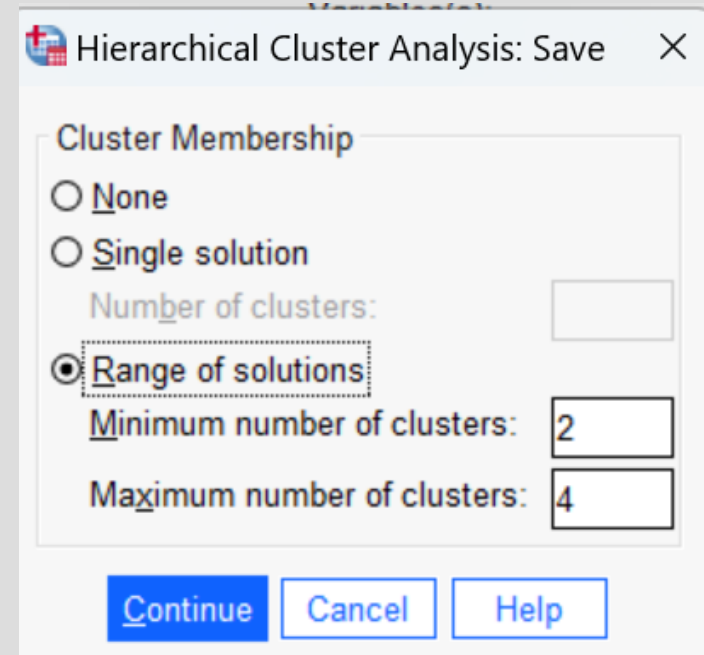
☐ Rescale to 0-1 range

Continue Cancel Help

# Υλοποίηση Hierarchical Clustering στο SPSS

## – Save

- Πατώντας το κουμπί επιλογών **Save** επιλέγουμε τις καινούριες μεταβλητές που θέλουμε να δημιουργήσουμε.



# Υλοποίηση Hierarchical Clustering στο SPSS

## – Save

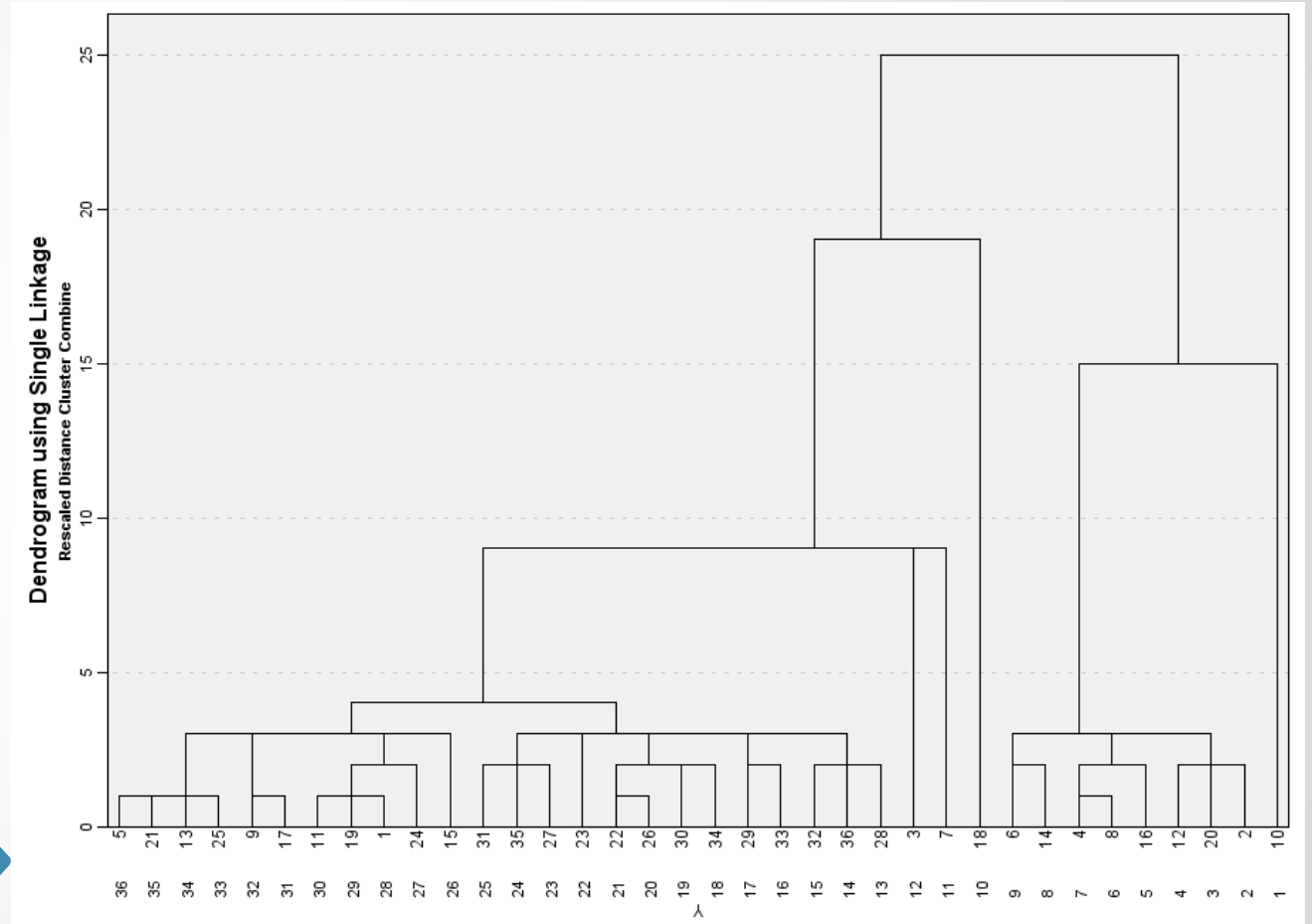
- Η επιλογή **Cluster membership** θα μας δημιουργήσει μια καινούρια στήλη όπου σε κάθε παρατήρηση θα δίνεται η τιμή της ομάδας που την κατατάξαμε.

# Υλοποίηση Hierarchical Clustering στο SPSS

## – Save

- Εδώ υπάρχει δυνατότητα μιας μοναδικής λύσης σε αριθμό clusters ή εύρους λύσεων.

# Παράδειγμα 2 – Υλοποίηση Αλγόριθμου Hierarchical Clustering στο SPSS – Αποτελέσματα Output



Δ.Β Παναγιωτάκος, Γ. Κουρλαμπά  
Εγχειρίδιο Χειρισμού του Στατιστικού Προγράμματος SPSS

ΣΤ. Δημητριάδης, Στατιστική Επιχειρήσεων με Εφαρμογές σε SPSS και LISREL, Εκδόσεις Κριτική, 2016

Verykios, V., Kagklis, V., & Stavropoulos, E. 2015. *Η επιστήμη των δεδομένων μέσα από τη γλώσσα R* [Undergraduate textbook]. Kallipos, Open Academic Editions.  
<https://hdl.handle.net/11419/2972>

Χαλικιάς Μ. Ποσοτική Ανάλυση και Στοιχεία Θεωρίας Αποφάσεων στη Διοίκηση και Οικονομία με Χρήση Λογισμικών EXCEL, ISALOS και SPSS, Εκδόσεις Broken Hill Publishers Ltd, 2021



## Βιβλιογραφία