

Εισαγωγή στην Τεχνητή Νοημοσύνη

(ΜΠΔ 306)

Εργαστήριο

Εργαστηριακός βοηθός: Κατερίνα Γιαννακάκη

e-mail: agiannakaki1@isc.tuc.gr

Χειμερινό εξάμηνο 2020-2021

Clustering Συσταδοποίηση

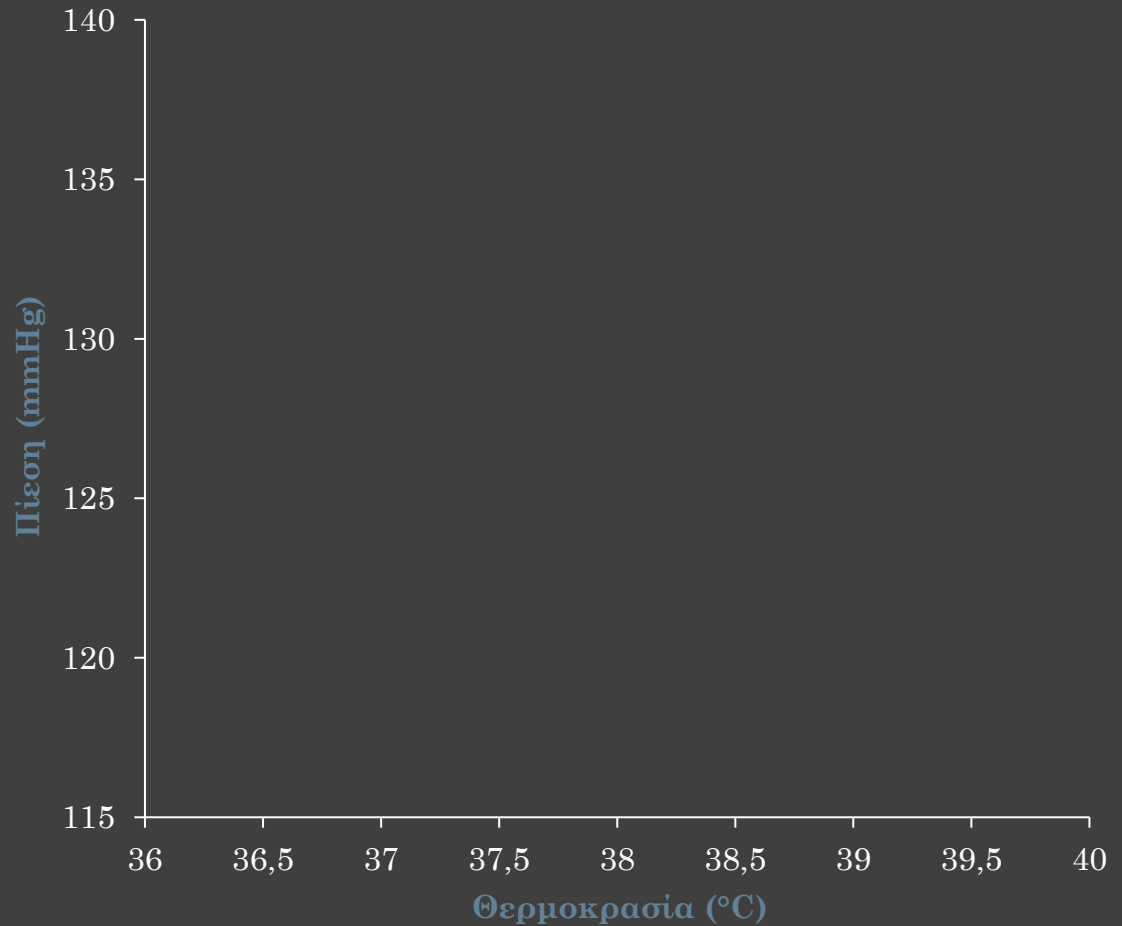
Εργαστήριο 5ο

Παράδειγμα

Θερμοκρασία (°C)	Πίεση (mmHg)
36,6	118
38,2	132
36,4	121
37,6	129
38,9	136
36,5	122
36,4	120
39,1	139
38,8	137
37,9	131
36,7	119

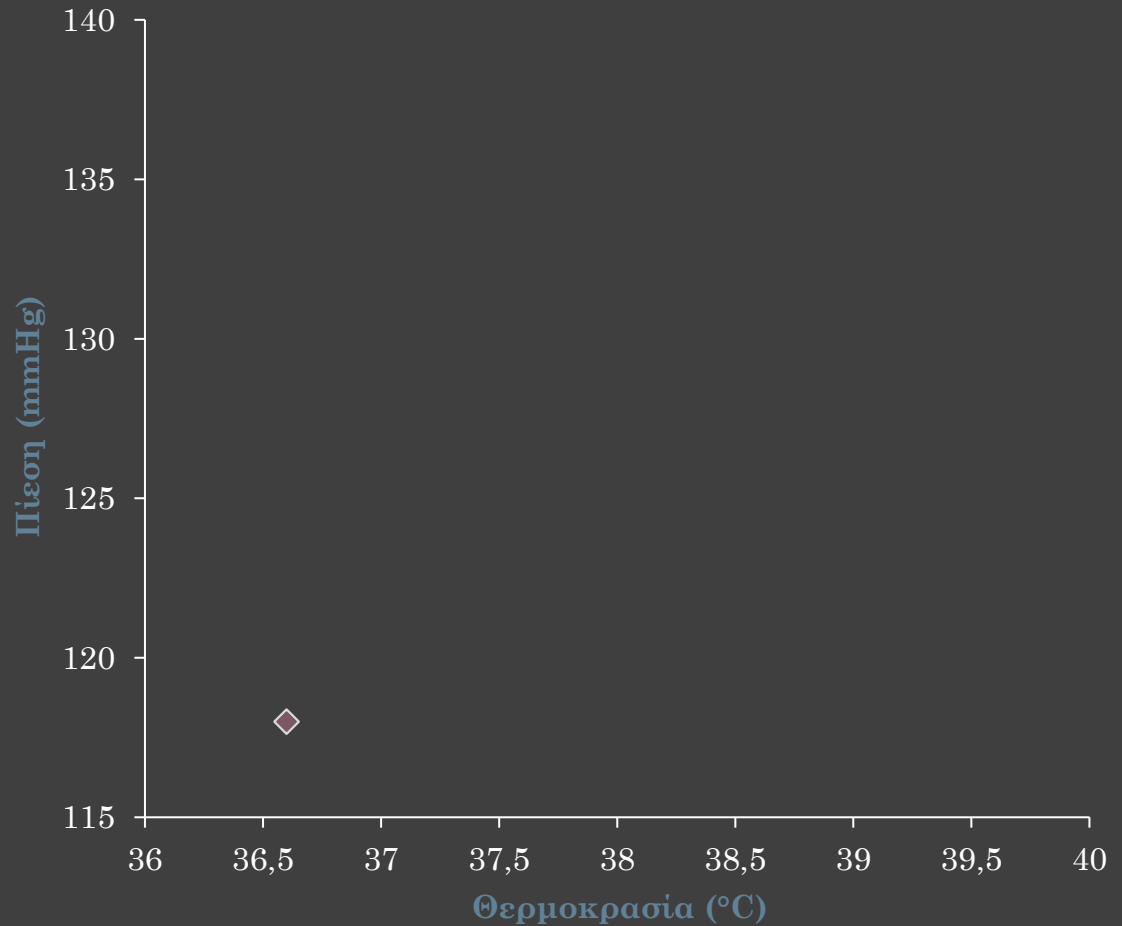
Παράδειγμα

Θερμοκρασία (°C)	Πίεση (mmHg)
36,6	118
38,2	132
36,4	121
37,6	129
38,9	136
36,5	122
36,4	120
39,1	139
38,8	137
37,9	131
36,7	119



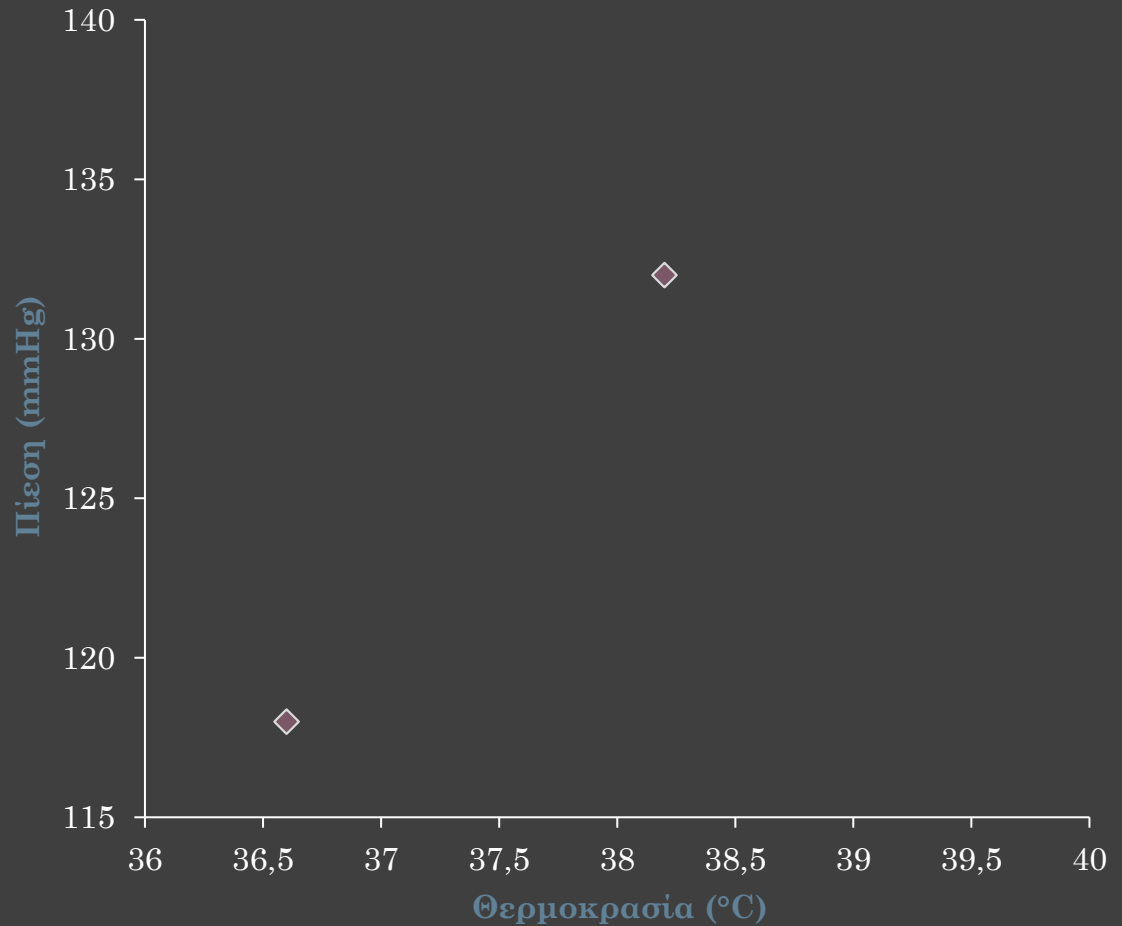
Παράδειγμα

Θερμοκρασία (°C)	Πίεση (mmHg)
36,6	118
38,2	132
36,4	121
37,6	129
38,9	136
36,5	122
36,4	120
39,1	139
38,8	137
37,9	131
36,7	119



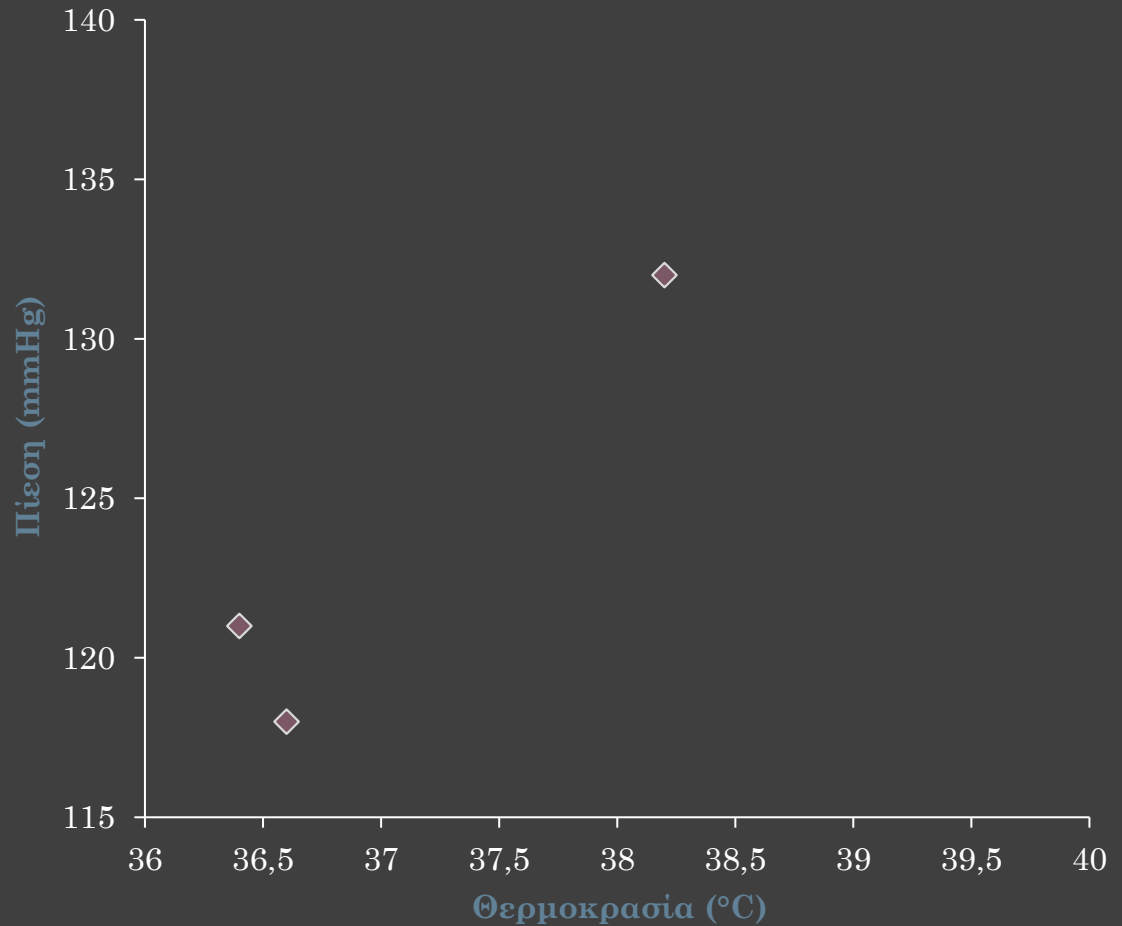
Παράδειγμα

Θερμοκρασία (°C)	Πίεση (mmHg)
36,6	118
38,2	132
36,4	121
37,6	129
38,9	136
36,5	122
36,4	120
39,1	139
38,8	137
37,9	131
36,7	119



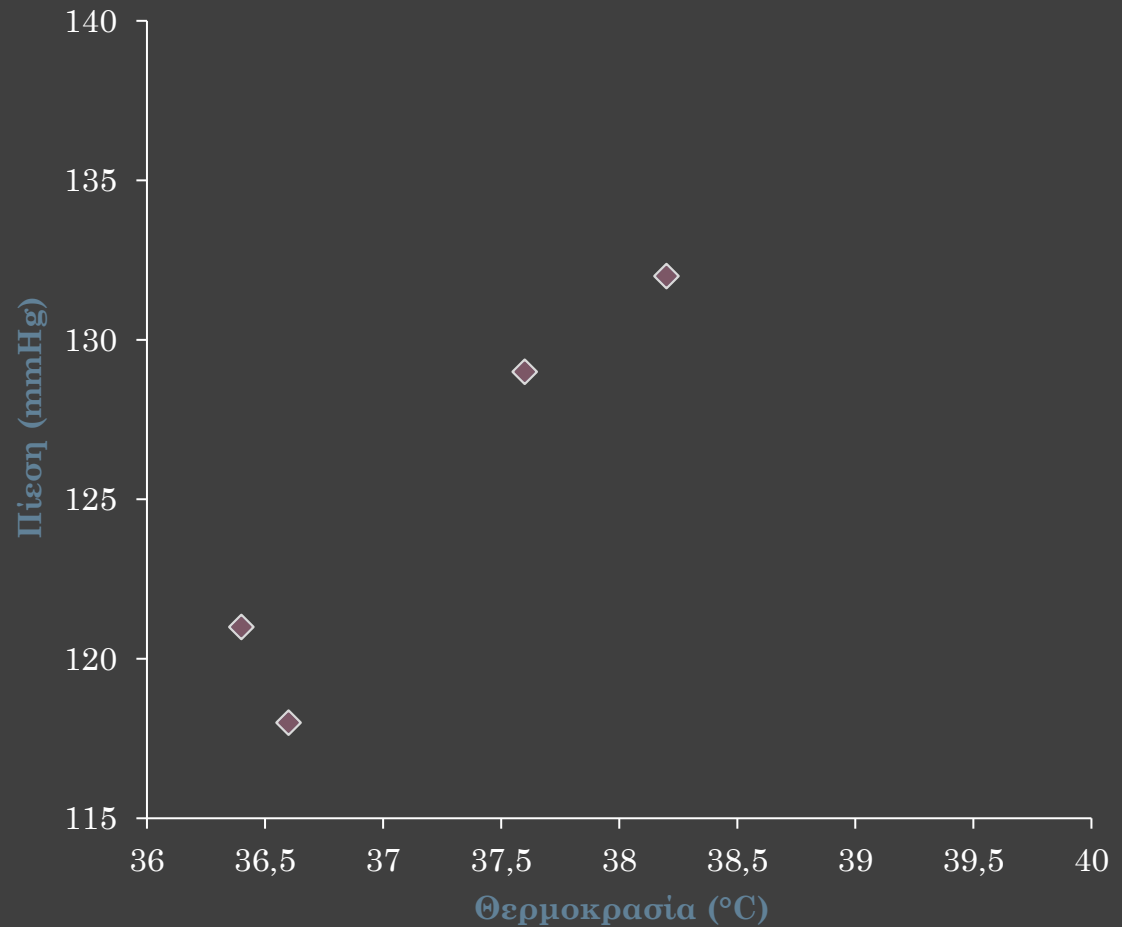
Παράδειγμα

Θερμοκρασία (°C)	Πίεση (mmHg)
36,6	118
38,2	132
36,4	121
37,6	129
38,9	136
36,5	122
36,4	120
39,1	139
38,8	137
37,9	131
36,7	119



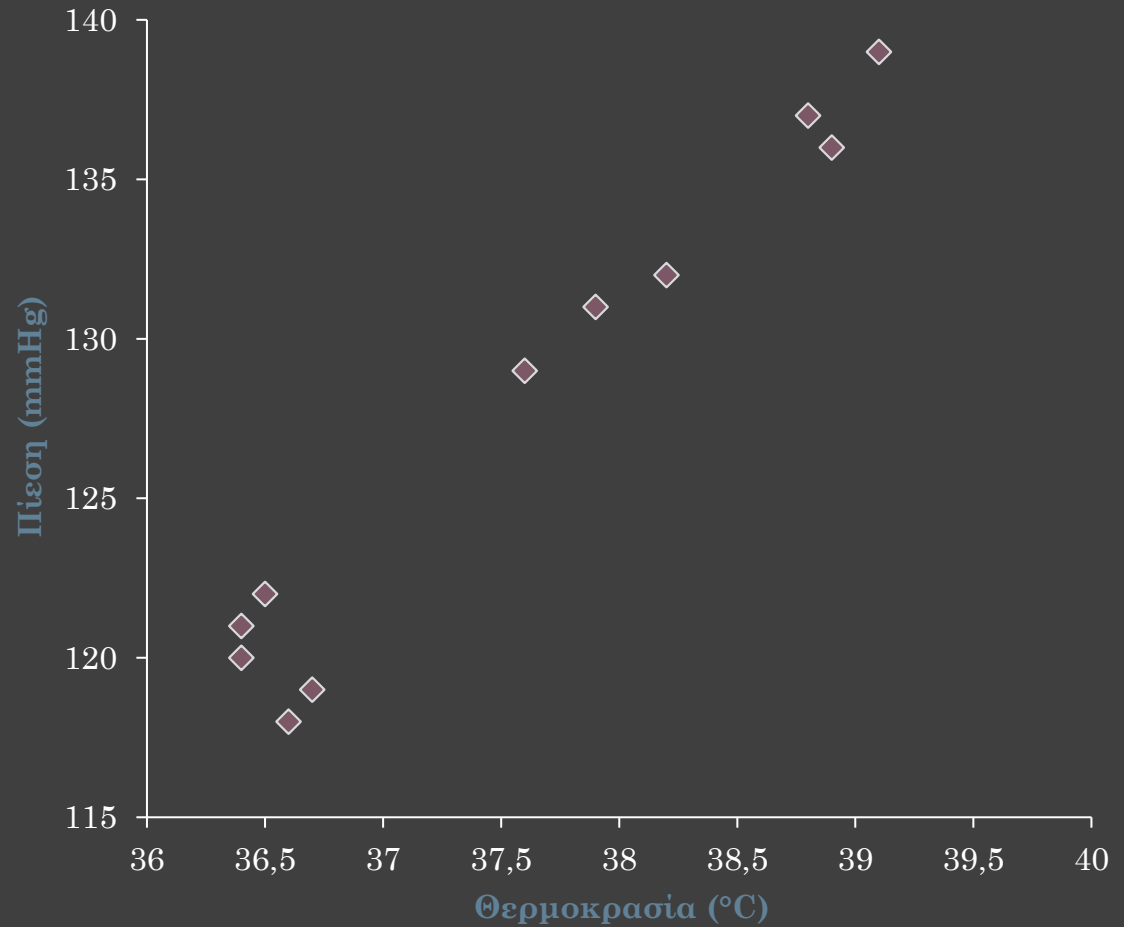
Παράδειγμα

Θερμοκρασία (°C)	Πίεση (mmHg)
36,6	118
38,2	132
36,4	121
37,6	129
38,9	136
36,5	122
36,4	120
39,1	139
38,8	137
37,9	131
36,7	119



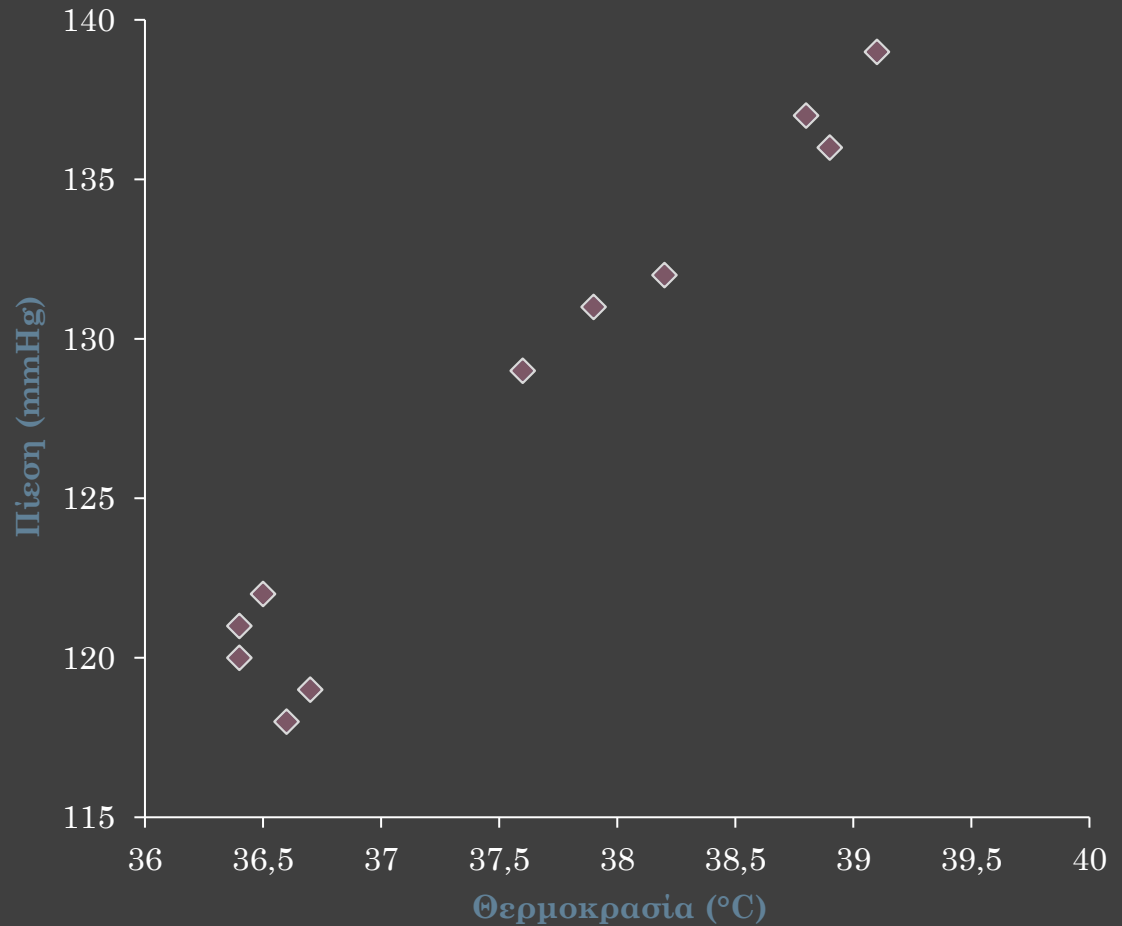
Παράδειγμα

Θερμοκρασία (°C)	Πίεση (mmHg)
36,6	118
38,2	132
36,4	121
37,6	129
38,9	136
36,5	122
36,4	120
39,1	139
38,8	137
37,9	131
36,7	119



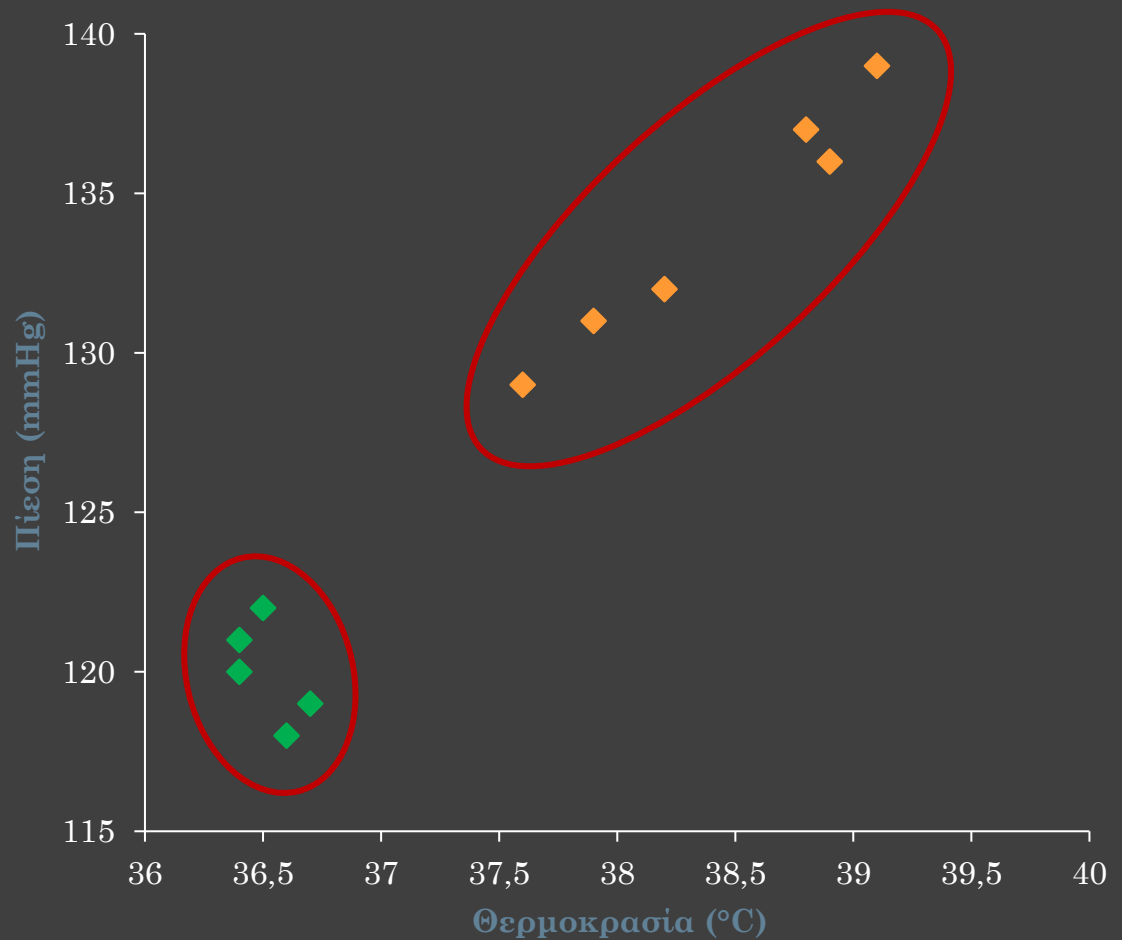
Παράδειγμα

Θερμοκρασία (°C)	Πίεση (mmHg)	Διάγνωση
36,6	118	Υγιής
38,2	132	Ασθενής
36,4	121	Υγιής
37,6	129	Ασθενής
38,9	136	Ασθενής
36,5	122	Υγιής
36,4	120	Υγιής
39,1	139	Ασθενής
38,8	137	Ασθενής
37,9	131	Ασθενής
36,7	119	Υγιής



Παράδειγμα

Θερμοκρασία (°C)	Πίεση (mmHg)	Διάγνωση
36,6	118	Υγιής
38,2	132	Ασθενής
36,4	121	Υγιής
37,6	129	Ασθενής
38,9	136	Ασθενής
36,5	122	Υγιής
36,4	120	Υγιής
39,1	139	Ασθενής
38,8	137	Ασθενής
37,9	131	Ασθενής
36,7	119	Υγιής



Συσταδοποίηση (Clustering)

Διαχωρισμός ενός συνόλου δεδομένων σε ομάδες, με τέτοιο τρόπο ώστε να ικανοποιούνται **ταυτόχρονα** τα δυο παρακάτω κριτήρια:

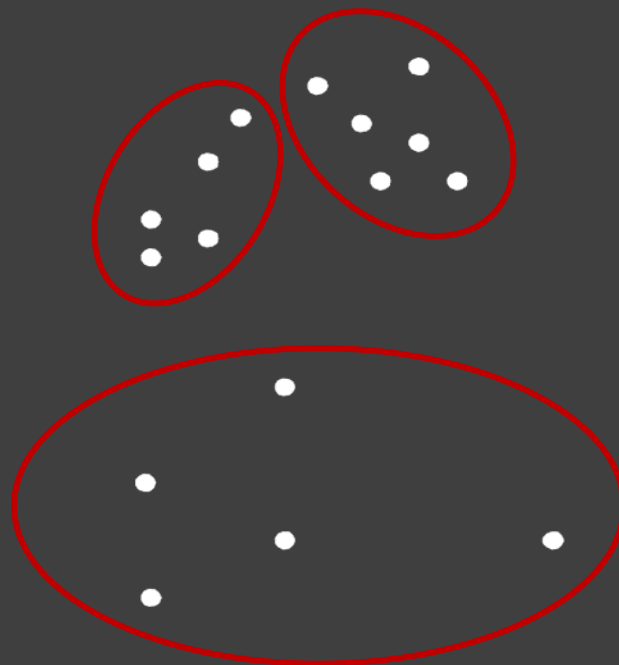
στην ίδια ομάδα να ανήκουν σημεία που:

- μοιάζουν όσο το δυνατόν περισσότερο μεταξύ τους και
- διαφέρουν όσο το δυνατόν περισσότερο με τα σημεία που ανήκουν σε άλλες ομάδες

Είδη Συσταδοποίησης

✓ Διαχωριστική Συσταδοποίηση (Partitional Clustering)

Διαμερισμός των αντικειμένων σε μη επικαλυπτόμενα υποσύνολα (συστάδες): κάθε αντικείμενο ανήκει σε ακριβώς ένα υποσύνολο.



διαχωριστική συσταδοποίηση

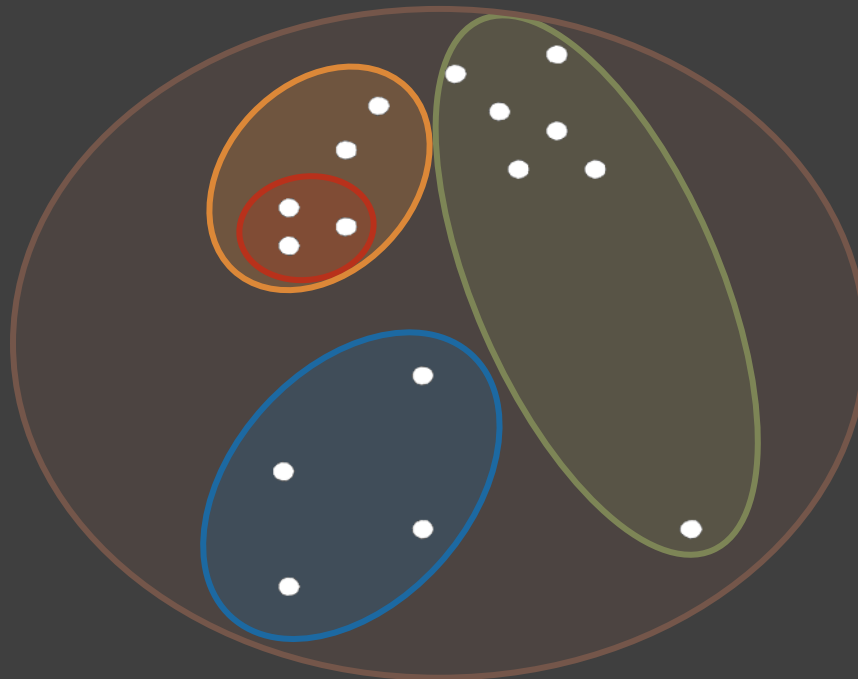
Είδη Συσταδοποίησης

- ✓ **Διαχωριστική Συσταδοποίηση (Partitional Clustering)**

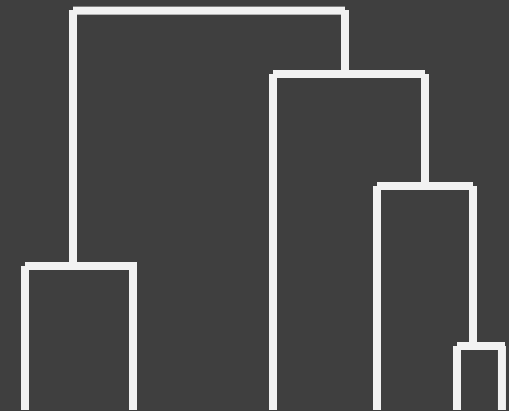
Διαμερισμός των αντικειμένων σε μη επικαλυπτόμενα υποσύνολα (συστάδες): κάθε αντικείμενο ανήκει σε ακριβώς ένα υποσύνολο.

- ✓ **Ιεραρχική Συσταδοποίηση (Hierarchical clustering)**

Διαμερισμός των αντικειμένων σε εμφωλευμένες (nested) συστάδες: μια συστάδα έχει υπο-συστάδες οργανωμένες σε ένα ιεραρχικό δέντρο

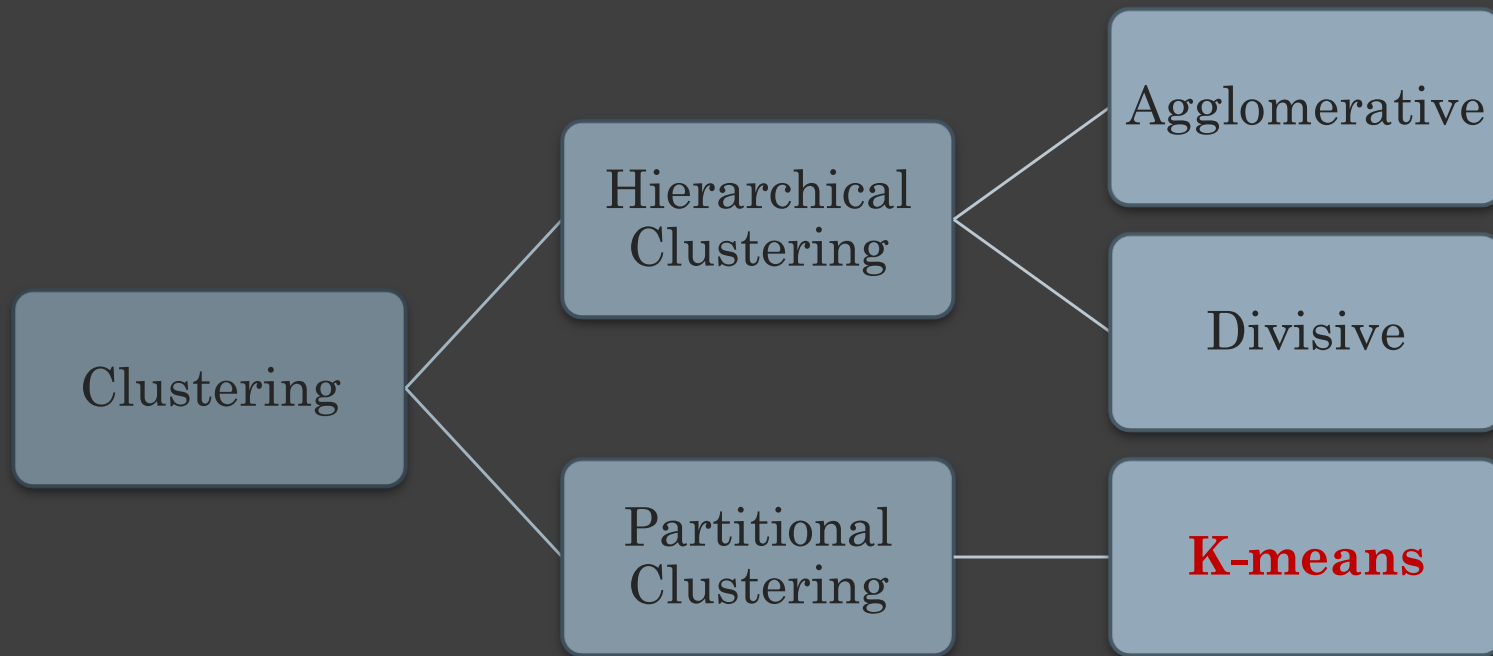


ιεραρχική συσταδοποίηση



δενδρόγραμμα

Clustering



K-means

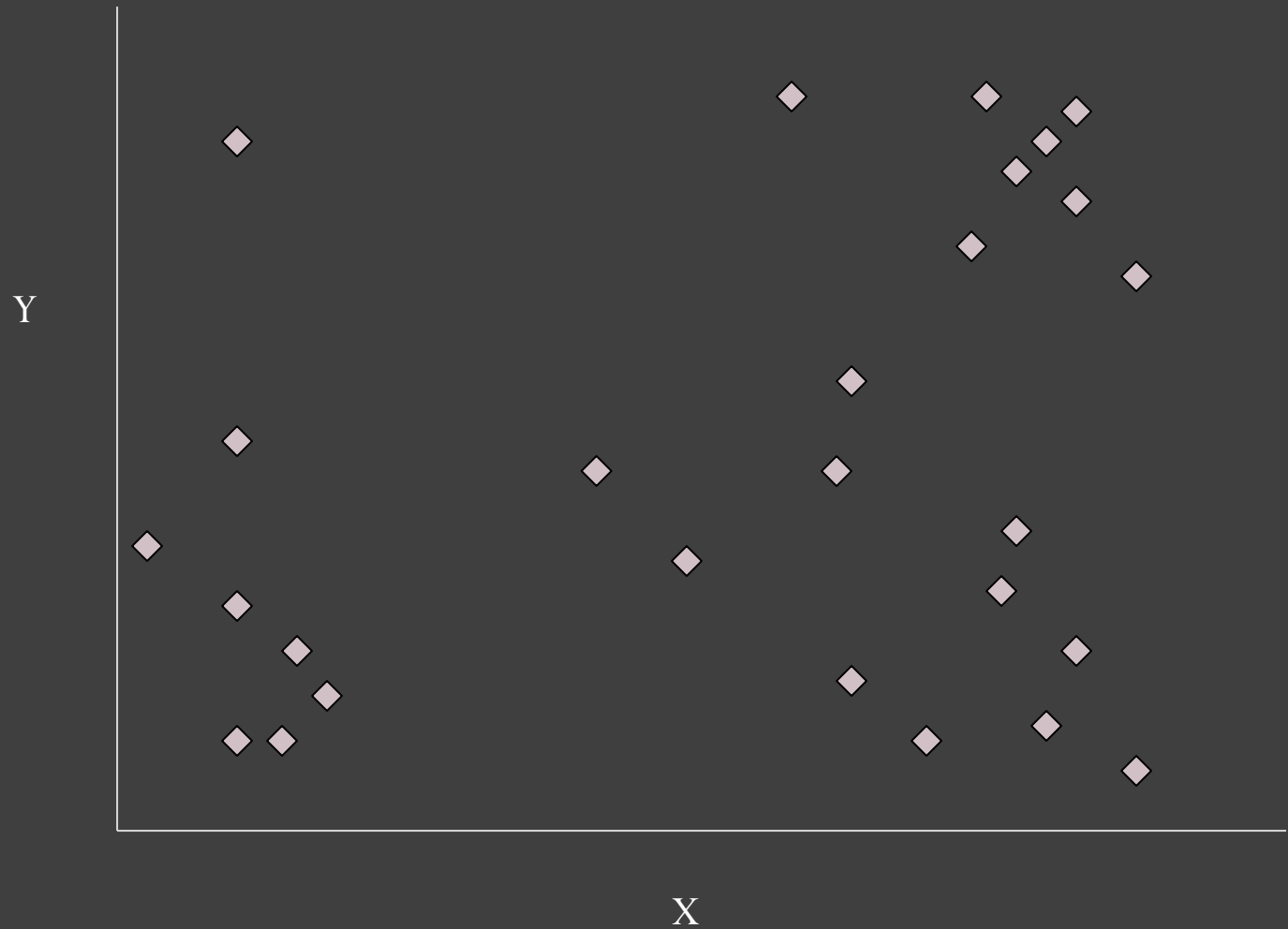
- Διαχωριστικός αλγόριθμος
- Κάθε συστάδα συσχετίζεται με ένα κεντρικό σημείο (centroid)
- Κάθε σημείο ανατίθεται στη συστάδα με το κοντινότερο κεντρικό σημείο
- Ο αριθμός των ομάδων (K) είναι η είσοδος στον αλγόριθμο

K-means

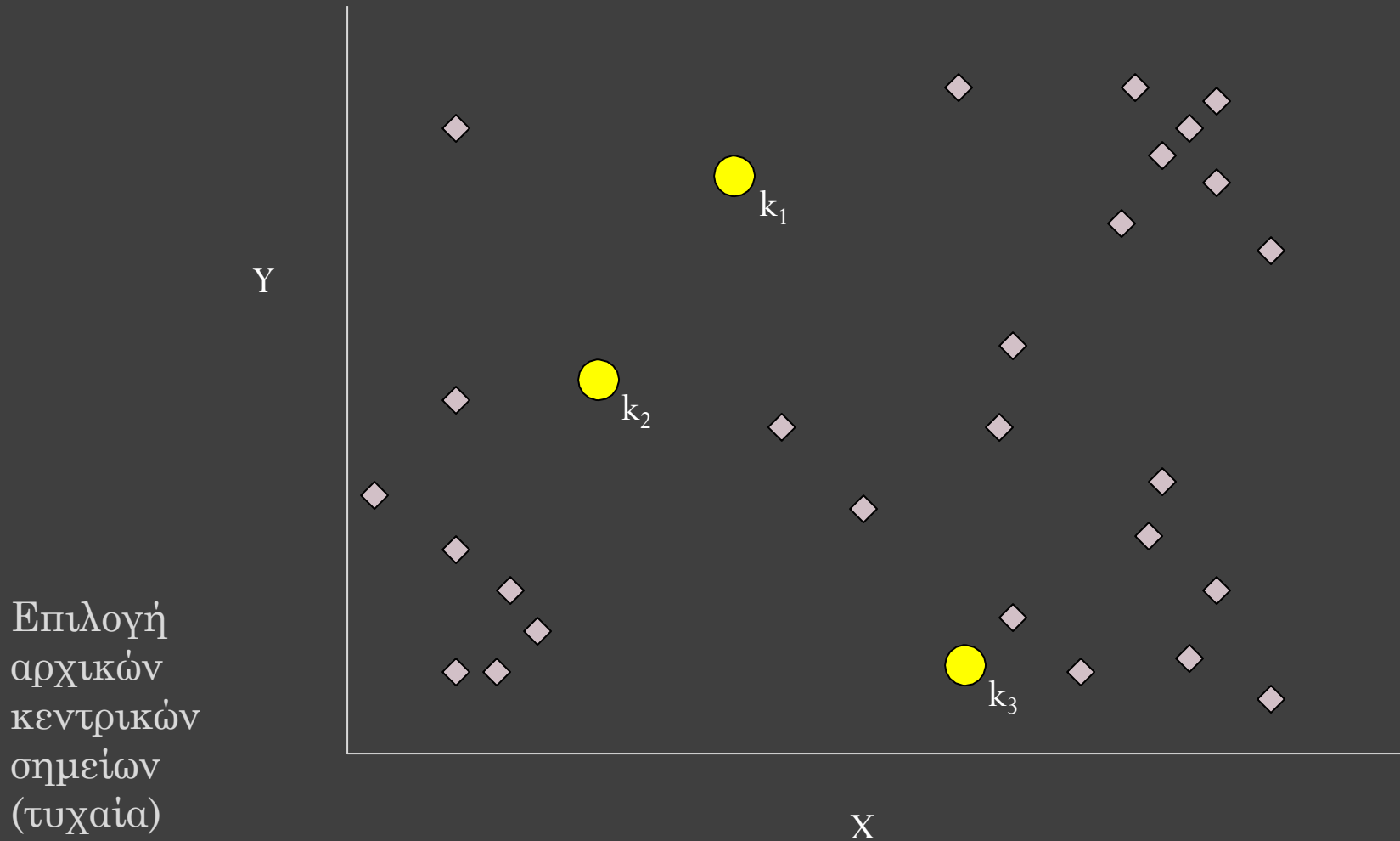
Βασικός αλγόριθμος

- Επιλογή K σημείων ως αρχικά κεντρικά σημεία
- Εκτέλεση των παρακάτω βημάτων επαναληπτικά μέχρι τα κεντρικά σημεία να μην αλλάζουν:
 - Ανάθεση όλων των αρχικών σημείων στο κοντινότερο τους από τα K κεντρικά σημεία
 - Επανα-υπολογισμός του κεντρικού σημείου κάθε συστάδας

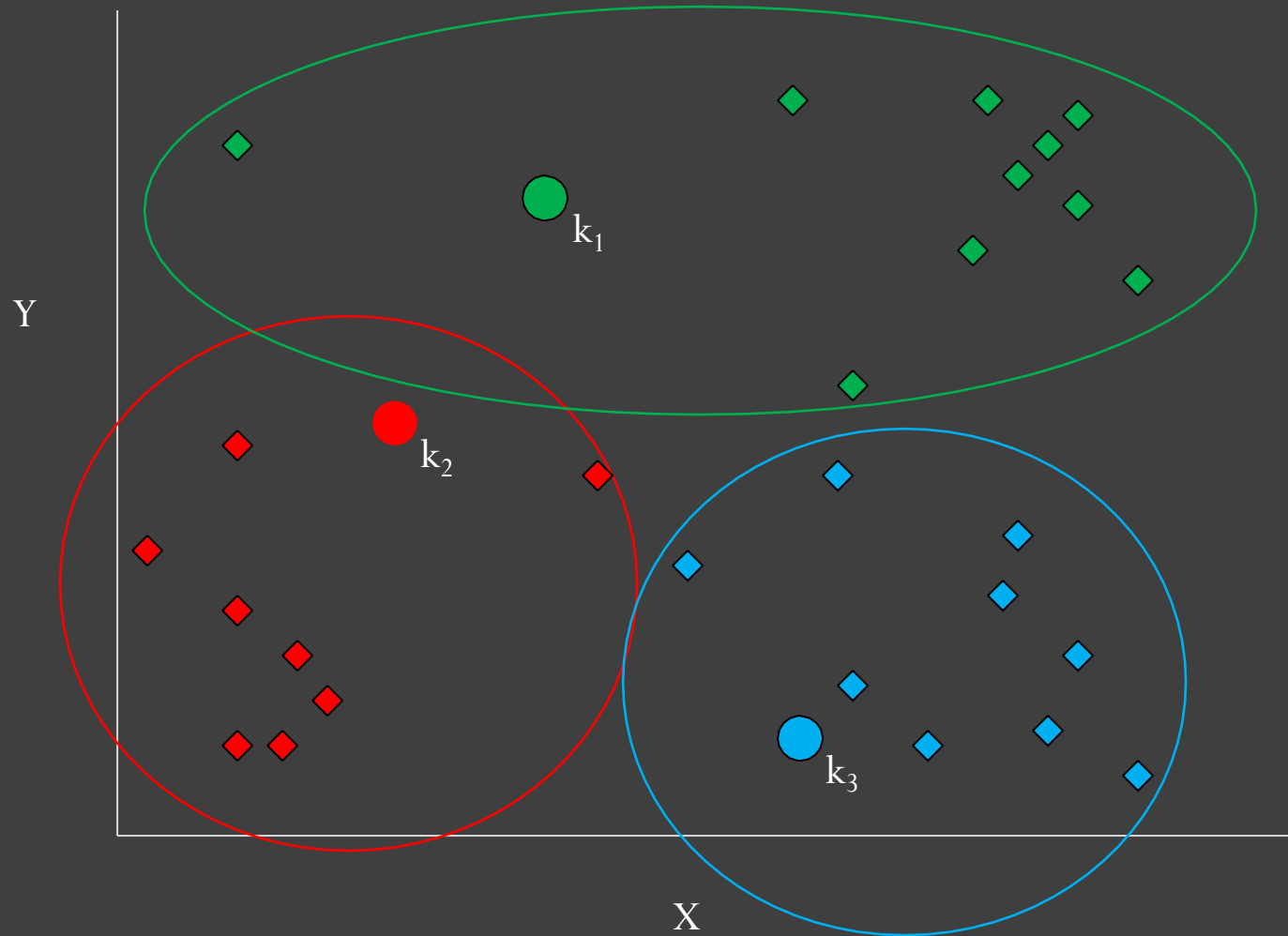
K-means



K-means



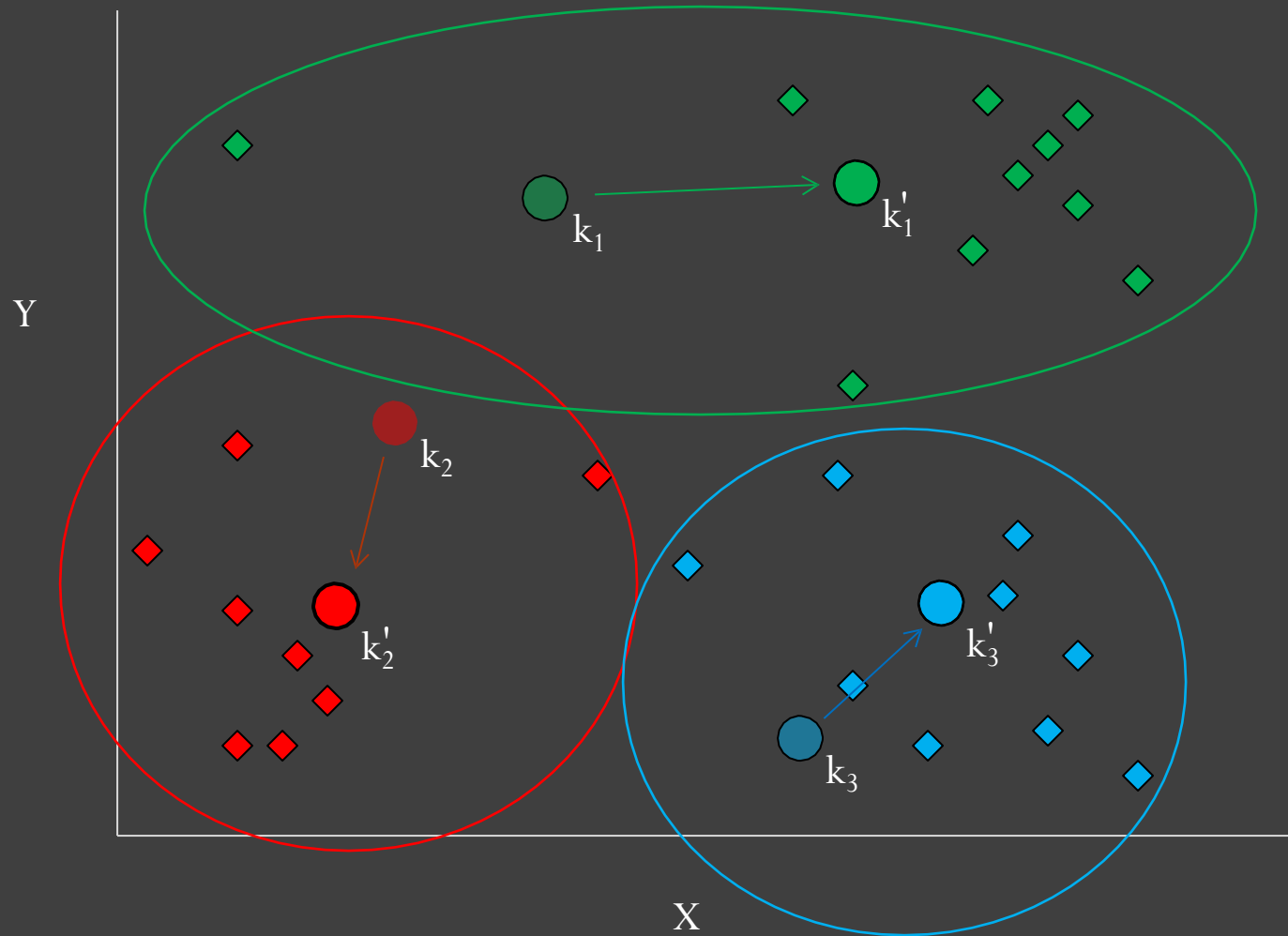
K-means



Ανάθεση κάθε
σημείου στο
κοντινότερο
κεντρικό
σημείο

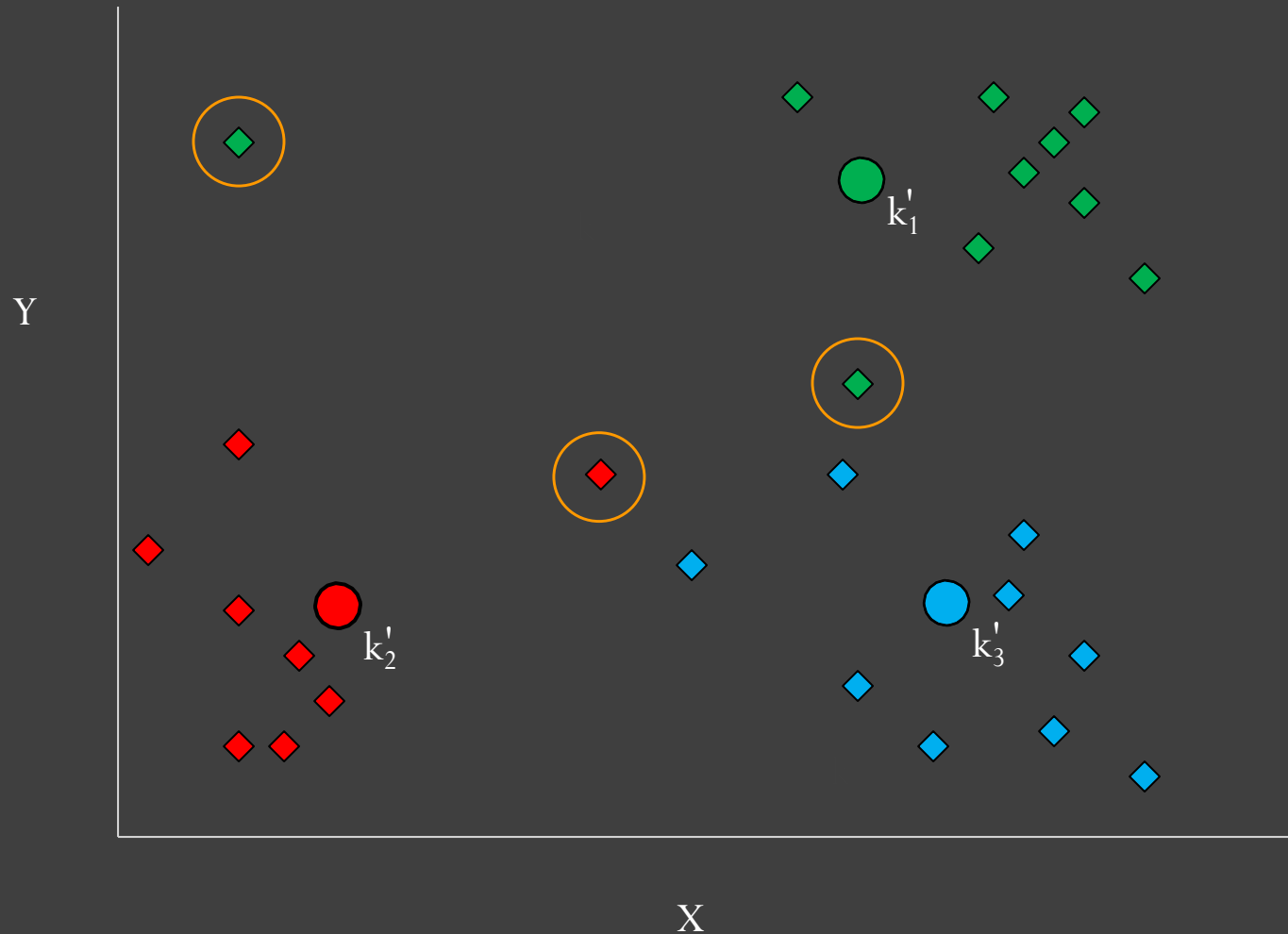
K-means

Επαν-
υπολογισμός
κεντρικών
σημείων:
κέντρο βάρους
ομάδας



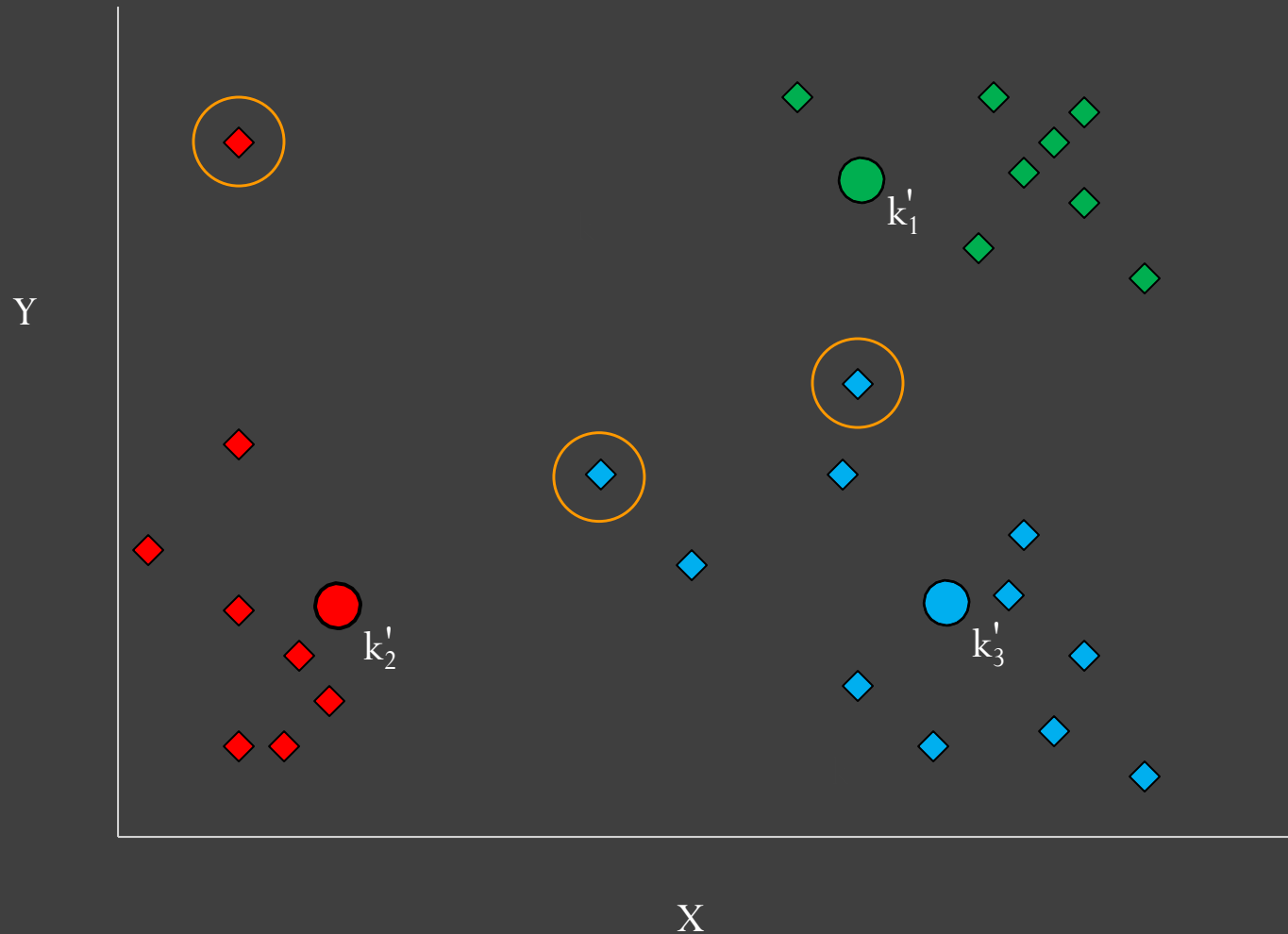
K-means

Επανά-
ανάθεση των
σημείων στα
κοντινότερα
κεντρικά
σημεία

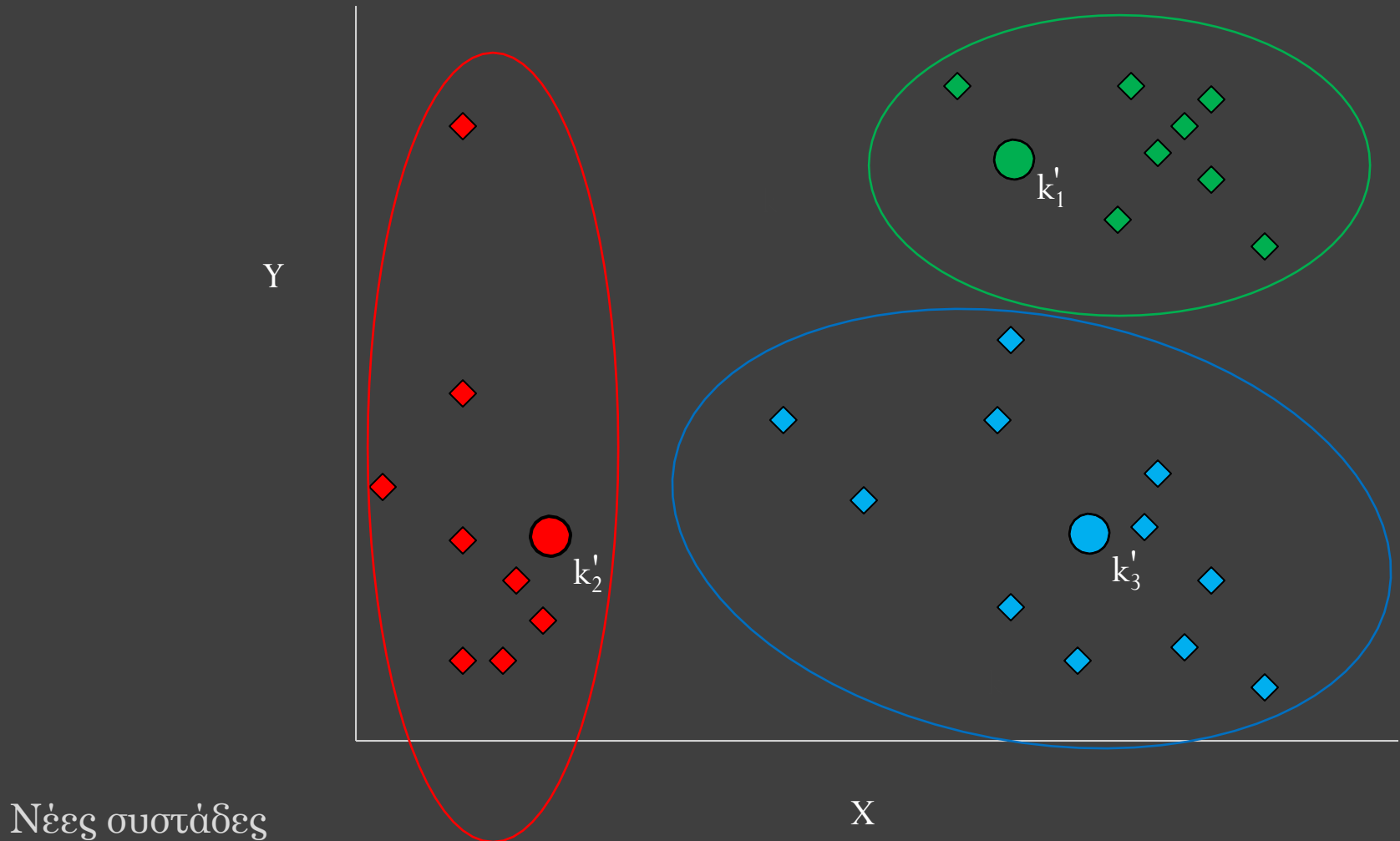


K-means

Επανά-
ανάθεση των
σημείων στα
κοντινότερα
κεντρικά
σημεία

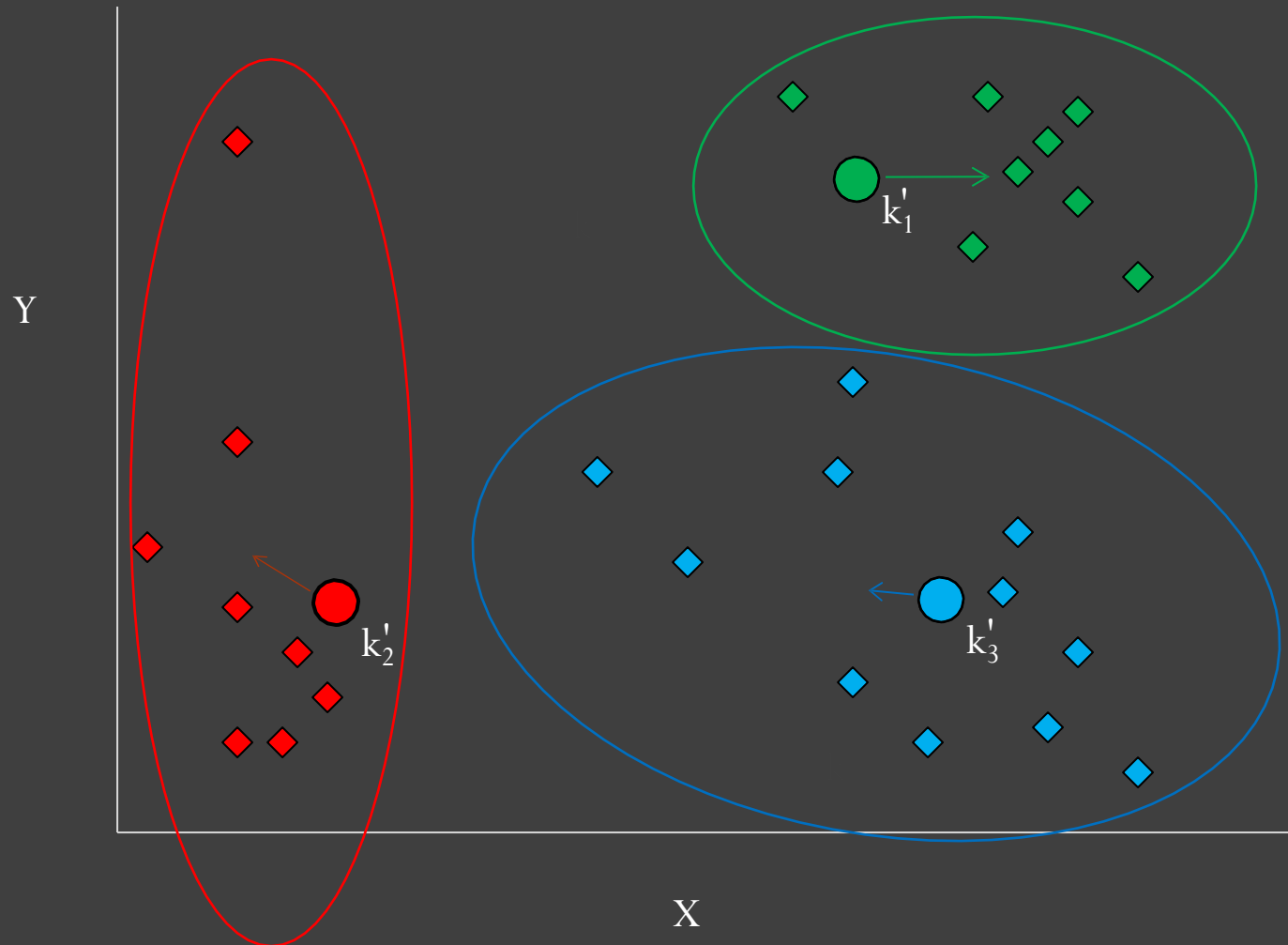


K-means



K-means

Επαν-
υπολογισμός
κεντρικών
σημείων
μέχρι κανένα
σημείο να μην
αλλάζει ομάδα



K-means

- ❖ Τα αρχικά κεντρικά σημεία συνήθως επιλέγονται τυχαία
- ❖ Πρέπει να ορίσουμε εμείς τον αριθμό των αρχικών κεντρικών σημείων και άρα των συστάδων που θα δημιουργηθούν
- ❖ Οι συστάδες που δημιουργούνται διαφέρουν σε κάθε βήμα του αλγορίθμου.
- ❖ Η εγγύτητα των σημείων υπολογίζεται με βάση κάποια απόσταση
- ❖ Επειδή η απόσταση υπολογίζεται συχνά πρέπει να είναι σχετικά απλή, συχνά χρησιμοποιείται η Ευκλείδεια απόσταση
- ❖ Το κεντρικό σημείο είναι συνήθως το μέσον των σημείων της συστάδας (το οποίο συνήθως δεν είναι ένα από τα δεδομένα)

Matlab

kmeans ()

- ✓ Εφαρμόζει τον αλγόριθμο K-means
- ✓ Παίρνει διάφορα ορίσματα (μέσα στην παρένθεση)
- ✓ Επιστρέφει διάφορους πίνακες/διανύσματα

kmeans

k-means clustering

Syntax

```
idx = kmeans(X,k)
idx = kmeans(X,k,Name,Value)
[idx,C] = kmeans( __ )
[idx,C,sumd] = kmeans( __ )
[idx,C,sumd,D] = kmeans( __ )
```

αναλυτικά: <https://www.mathworks.com/help/stats/kmeans.html>

Matlab

```
idx = kmeans(X,k)
```

Εφαρμόζει τον αλγόριθμο k-means για να δημιουργήσει **k** συστάδες από τα δεδομένα που βρίσκονται στον πίνακα **X** (διαστάσεων $n \times p$)

Επιστρέφει ένα διάνυσμα **idx** (διαστάσεων $n \times 1$) με τις συστάδες στις οποίες ανήκουν τα σημεία των αρχικών δεδομένων. Κάθε στοιχείο του **idx** αντιστοιχεί σε μια γραμμή του **X**, και η τιμή κάθε στοιχείου του **idx** είναι η συστάδα στην οποία ανήκει το αντίστοιχο σημείο. Π.χ.:

```
idx' = [2 2 1 1 1 3 3 3 3 3 3 2 2 2 1 1 1 1 1 3 3]
```

Ο k-means έχει αναθέσει το πρώτο και το δεύτερο σημείο στη συστάδα 2, το τρίτο, τέταρτο, πέμπτο στη συστάδα 1, το 6^ο-12^ο στη συστάδα 3, κ.ο.κ.

Matlab

`kmeans()`

Δεδομένα: πίνακας (έστω X)

γραμμές: καταχωρήσεις

στήλες: χαρακτηριστικά
(διαστάσεις του προβλήματος)

	Θερμοκρασία (°C)	Πίεση (mmHg)	Διάγνωση
ασθενής #1	36,6	118	Υγιής
ασθενής #2	38,2	132	Ασθενής
ασθενής #3	36,4	121	Υγιής
ασθενής #4	37,6	129	Ασθενής
ασθενής #5	38,9	136	Ασθενής
ασθενής #6	36,5	122	Υγιής
ασθενής #7	36,4	120	Υγιής
ασθενής #8	39,1	139	Ασθενής
ασθενής #9	38,8	137	Ασθενής
ασθενής #10	37,9	131	Ασθενής
ασθενής #11	36,7	119	Υγιής

`idx = kmeans(X,2)`

`idx' = [1 2 1 2 2 1 1 2 2 2 1]`

Matlab: kmeans () - Παραμετροποίηση

Απόσταση

```
idx = kmeans(X,k,'distance','sqeuclidean');
```

Για να υπολογίσει την απόσταση μεταξύ των σημείων, χρησιμοποιεί την τετραγωνική Ευκλείδεια απόσταση:

$$\sum_{i=1}^n (x_i - y_i)^2$$

Η παράμετρος `distance` μπορεί να πάρει και άλλες τιμές:
'cityblock', 'cosine', 'correlation', 'hamming'

αναλυτικά:

https://www.mathworks.com/help/stats/kmeans.html#inputarg_Distance

Matlab: `kmeans` () - Παραμετροποίηση

Εμφάνιση βημάτων - πληροφοριών

```
idx =  
kmeans(X,k,'distance','sqeuclidean','display','iter');
```

Εμφανίζει πληροφορίες για κάθε βήμα του αλγορίθμου που υλοποιεί.

Η παράμετρος `display` μπορεί να πάρει και άλλες τιμές:

`'final'` : εμφανίζει τα αποτελέσματα του τελευταίου βήματος
`'off'` : δεν εμφανίζει τίποτα

αναλυτικά:

https://www.mathworks.com/help/stats/kmeans.html#inputarg_Display

Matlab: kmeans ()

DOCUMENTATION

k-means clustering:

<https://www.mathworks.com/help/stats/kmeans.html#buefthh-3>

function kmeans () :

<https://www.mathworks.com/help/stats/kmeans.html>

Matlab: kmeans () - Παράδειγμα

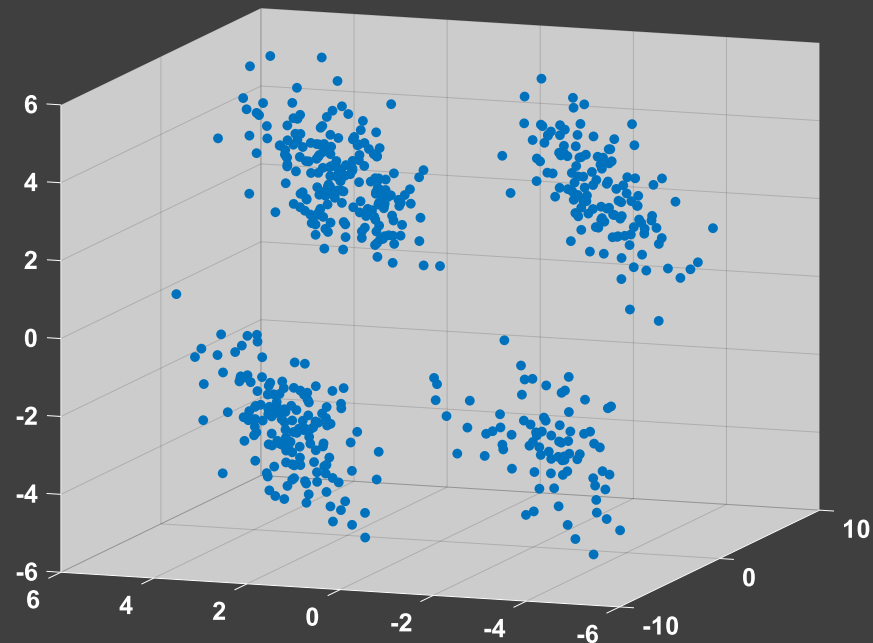
```
clear all; close all; clc;  
  
% load data  
load kmeansdata  
  
% display data  
figure;  
scatter3(X(:,1),X(:,2),X(:,3),'filled');
```

Σχόλια:

Ο πίνακας X έχει 4 στήλες, άρα το πρόβλημα έχει 4 διαστάσεις

Όμως μπορούμε να δούμε γραφική παράσταση σε μέχρι 3 διαστάσεις

Οπότε επιλέγουμε να χρησιμοποιήσουμε τις 3 πρώτες στήλες του X για να οπτικοποιήσουμε τα δεδομένα



Matlab: kmeans () - Παράδειγμα

```
%% run k-means with parameters:  
% Number of Clusters : 4  
% Distance           : squared Euclidean  
% Display            : show results of each step  
idx = kmeans(X,4,'distance','sqeuclidean','display','iter');
```

Matlab: kmeans () - Παράδειγμα

```
%% run k-means with parameters:  
% Number of Clusters : 4  
% Distance : squared Euclidean  
% Display : show results of each step  
idx = kmeans(X, 4, 'distance', 'sqeuclidean', 'display', 'iter');
```

διάνυσμα που
περιέχει το
cluster στο οποίο
έχει ανατεθεί
κάθε καταχώρηση

πίνακας με
τα δεδομένα

ο αριθμός των
clusters που
θέλουμε

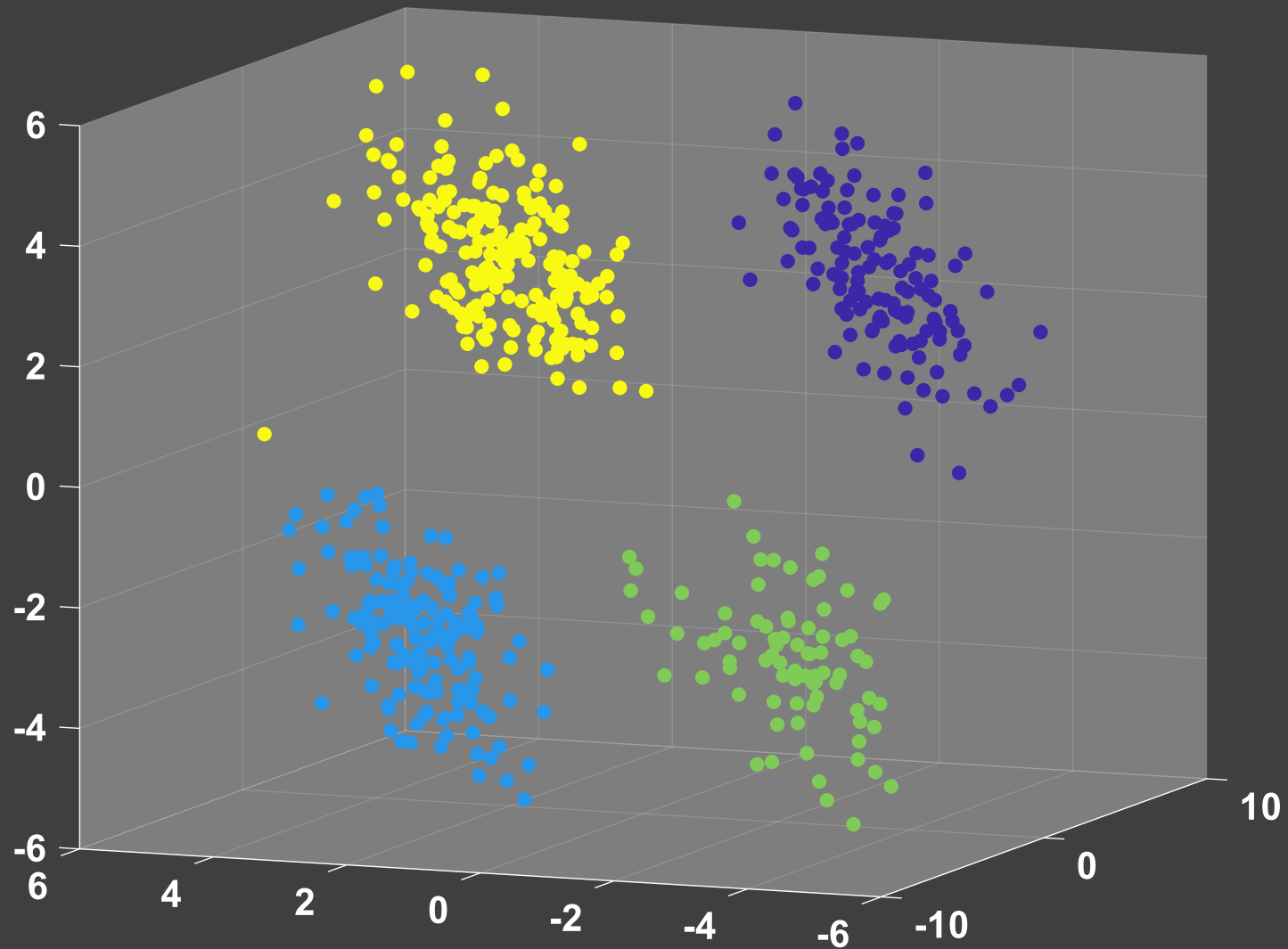
ποιος τύπος θα χρησιμοποιηθεί
για τον υπολογισμό της
απόστασης μεταξύ των σημείων

Matlab: kmeans () - Παράδειγμα

```
%% run k-means with parameters:
% Number of Clusters : 4
% Distance           : squared Euclidean
% Display            : show results of each step
idx = kmeans(X,4,'distance','sqeuclidean','display','iter');

% display clustered data
figure;
scatter3(X(:,1),X(:,2),X(:,3),[],idx,'filled');
```

Matlab: `kmeans()` - Παράδειγμα



Matlab: `kmeans()` - Παράδειγμα

Για την εμφάνιση της γραφικής παράστασης χρησιμοποιήθηκε η συνάρτηση `scatter3()`:

```
scatter3(X(:,1),X(:,2),X(:,3),'filled');  
scatter3(X(:,1),X(:,2),X(:,3),[],idx,'filled');
```

αναλυτικά:

<https://www.mathworks.com/help/matlab/ref/scatter3.html>

Τοπικά ελάχιστα

Η τελικές συστάδες που δίνει ο k-means επηρεάζονται από τα αρχικά σημεία που επιλέγονται τυχαία.

Αυτό σημαίνει ότι είναι πιθανό κάποια λύση να είναι **τοπικό βέλτιστο**. Για να μετριάσουμε αυτήν την πιθανότητα τρέχουμε τον αλγόριθμο περισσότερες από μια φορές ώστε κάθε φορά να έχουμε διαφορετικά σημεία εκκίνησης.

Matlab: kmeans () – 'replicates'

Η παράμετρος replicates

```
idx = kmeans(X,k,'distance','sqeuclidean', . . .  
           'display','final', . . .  
           'replicates',10);
```

Τρέχει τον αλγόριθμο 10 φορές, κάθε φορά με διαφορετικά αρχικά σημεία, και επιλέγει το καλύτερο αποτέλεσμα.

Η συνάρτηση Silhouette

- Μέθοδος αξιολόγησης της συνοχής εντός των clusters
- Παρέχει μια συνοπτική γραφική αναπαράσταση του πόσο καλά έχει ανατεθεί κάθε σημείο στο cluster του
- Ένα μέτρο ομοιότητας κάθε σημείου με το cluster του σε σύγκριση με τα υπόλοιπα clusters

$$s(i) = \begin{cases} 1 - a(i)/b(i), & \text{αν } a(i) < b(i) \\ 0, & \text{αν } a(i) = b(i) \\ b(i)/a(i) - 1, & \text{αν } a(i) > b(i) \end{cases}$$

$$-1 \leq s(i) \leq 1$$

$a(i)$: η μέση απόσταση του σημείου i από όλα τα σημεία του ίδιου cluster

$b(i)$: η μέση απόσταση του σημείου i από τα σημεία των υπόλοιπων clusters. Το cluster όπου το $b(i)$ είναι μικρότερο είναι το γειτονικό cluster

Η συνάρτηση Silhouette

Όσο πιο κοντά στο **1** είναι η τιμή του $s(i)$, τόσο πιο επιτυχημένη είναι η συσταδοποίηση, δηλαδή η ανάθεση του σημείου i στο cluster του

Όσο πιο κοντά στο **-1** είναι η τιμή του $s(i)$, τόσο πιο αποτυχημένη είναι η συσταδοποίηση, δηλαδή το σημείο i θα ταίριαζε καλύτερα σε άλλο cluster

Όταν η τιμή του $s(i)$ είναι κοντά στο **μηδέν**, τότε το σημείο i βρίσκεται στο όριο μεταξύ δυο clusters, δεν έχει μεγάλη διαφορά σε πιο από τα δυο θα ανατεθεί

Ο Συντελεστής Silhouette (silhouette coefficient)

Ο συντελεστής silhouette είναι ο μέσος όρος των τιμών silhouette από όλα τα σημεία του προβλήματος

Είναι ένα ποσοτικό μέγεθος αξιολόγησης της ποιότητας της συσταδοποίησης

Matlab: silhouette()

```
s = silhouette(X,idx,'sqEuclidean');
```

Υπολογίζει τις τιμές της συνάρτησης Silhouette για τα δεδομένα του πίνακα **X** (διαστάσεων $n \times p$) και το διάνυσμα των clusters **idx**.

Ο τύπος της απόστασης που θα χρησιμοποιήσει στον υπολογισμό των $a(i)$, $b(i)$ δίνεται από το τρίτο όρισμα. Μπορεί να πάρει και άλλες τιμές, π.χ. 'Euclidean', 'cityblock', 'cosine', κ.λπ.

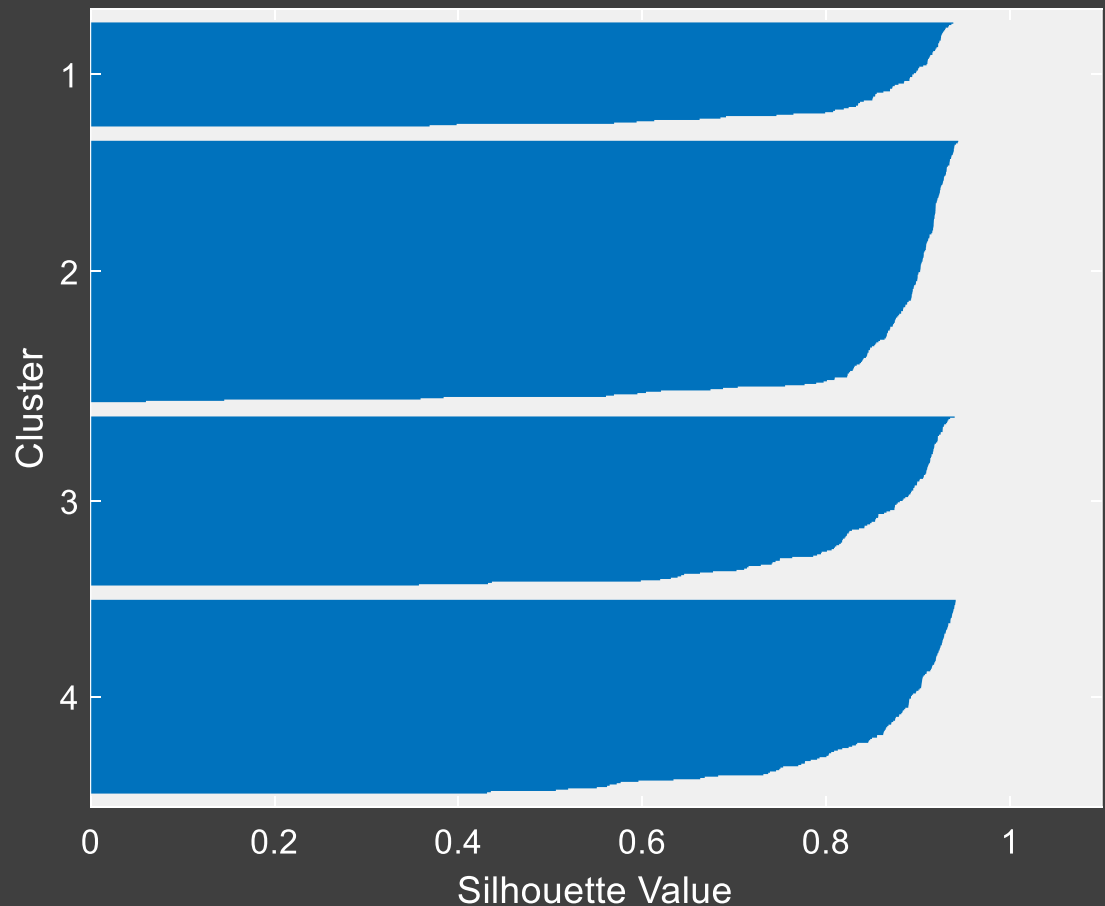
αναλυτικά: <https://www.mathworks.com/help/stats/silhouette.html>

Matlab: silhouette()

```
silhouette(X,idx,'sqEuclidean');
```

ΧΩΡΙΣ ΤΟ 's = '

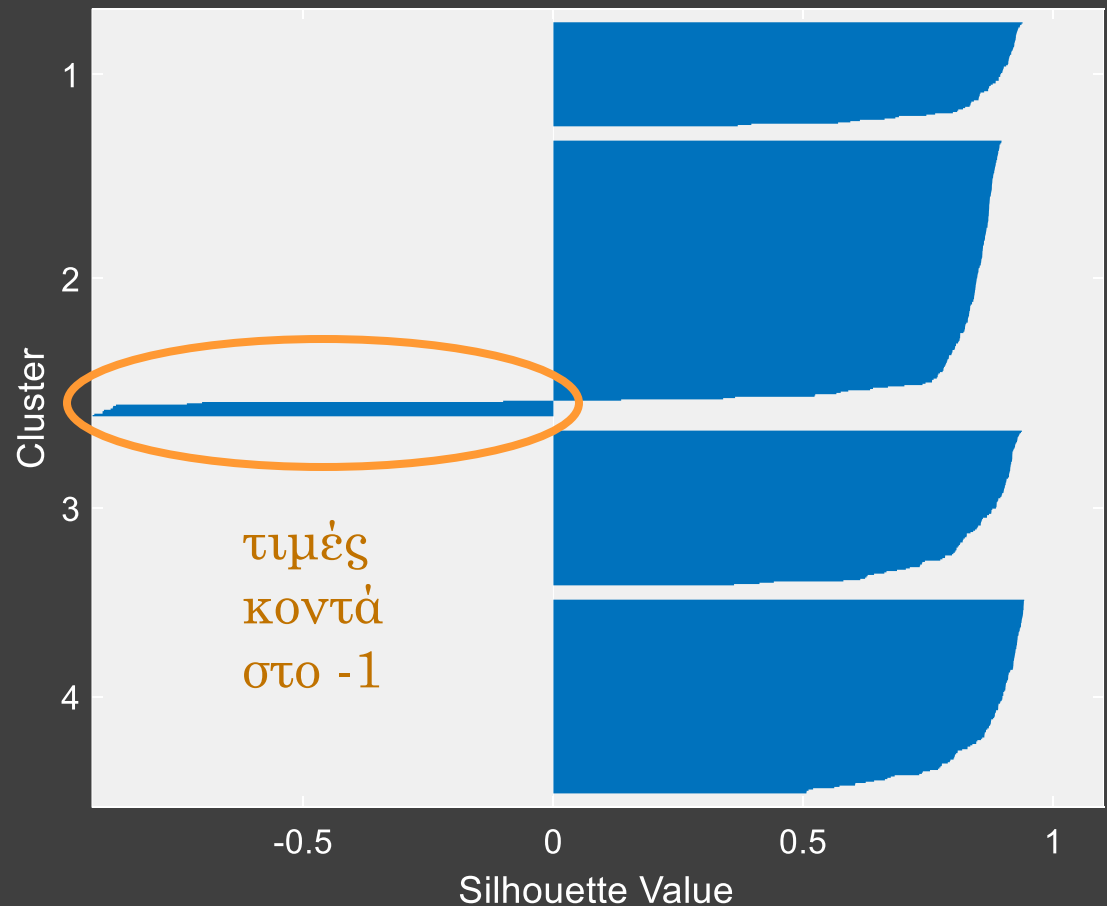
Εμφανίζει γράφημα με τις τιμές της silhouette για κάθε σημείο του κάθε cluster



Matlab: silhouette()

```
silhouette(X,idx2,'sqEuclidean');
```

Αν ο kmeans είχε
ομαδοποιήσει λάθος
κάποια δεδομένα στο
cluster 2



Matlab: silhouette ()

```
s = silhouette(X,idx,'sqEuclidean');
```

Άρα μπορώ να αξιολογήσω πόσο καλά έχει γίνει η συσταδοποίηση από το γράφημα της silhouette

Και να συγκρίνω διαφορετικά τρεξίματα (με διαφορετικές παραμέτρους, π.χ. για 3 και για 4 clusters) χρησιμοποιώντας τη μέση τιμή των τιμών της silhouette:

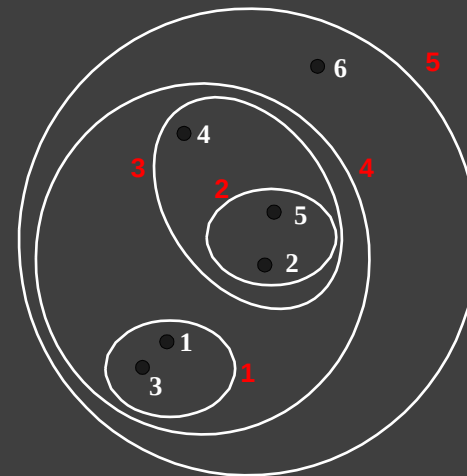
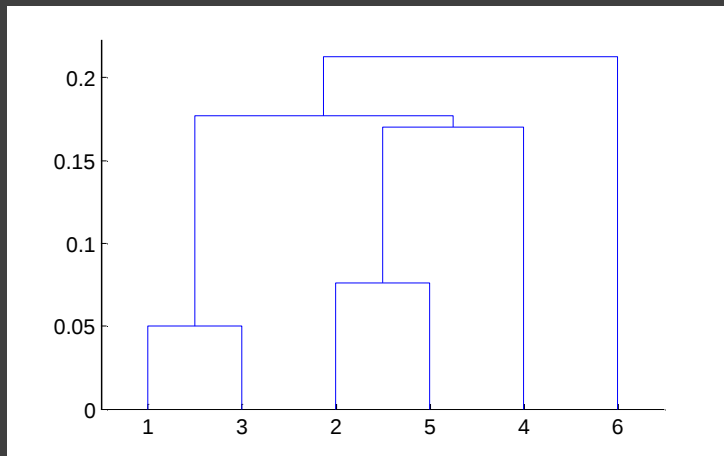
```
mean(s)
```

Όσο μεγαλύτερη η μέση τιμή, τόσο καλύτερη η συσταδοποίηση

Ιεραρχική Συσταδοποίηση

Ιεραρχική Συσταδοποίηση (Hierarchical clustering)

- Παράγει ένα σύνολο από εμφωλευμένες συστάδες οργανωμένες σε ένα ιεραρχικό δέντρο.
- Μπορεί να γίνει αναπαράσταση με ένα δενδρόγραμμα



Τύποι Ιεραρχικής Συσταδοποίησης

Συσσωρευτικός (Agglomerative)

Αρχίζει με το κάθε σημείο ως μια ξεχωριστή συστάδα

Σε κάθε βήμα συγχωνεύει το πιο κοντινό ζευγάρι συστάδων μέχρι να συγχωνευτούν όλες σε μία

Διαιρετικός (divisive)

Αρχίζει με μια συστάδα που περιέχει όλα τα σημεία

Σε κάθε βήμα διαχωρίζει μία συστάδα μέχρι κάθε συστάδα να περιέχει μόνο ένα σημείο

Ιεραρχική Συσταδοποίηση

- ✓ Δε χρειάζεται να υποθέσουμε ένα συγκεκριμένο αριθμό από συστάδες. Οποιοσδήποτε επιθυμητός αριθμός από συστάδες μπορεί να επιτευχθεί κόβοντας το δενδρόγραμμα στο κατάλληλο επίπεδο
- ✓ Διαχωρισμός ή συγχώνευση μιας ομάδας τη φορά
- ✓ Χρήση πίνακα ομοιότητας ή απόστασης

Ιεραρχική Συσταδοποίηση

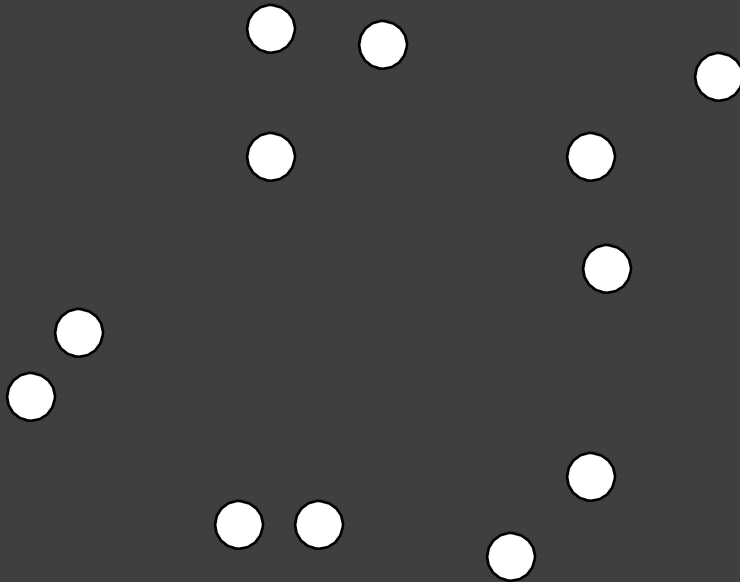
Βασικός αλγόριθμος

Έστω ότι αρχικά κάθε σημείο αποτελεί και μία συστάδα

- Υπολογισμός του πίνακα γειτνίασης
- Εκτέλεση των παρακάτω βημάτων επαναληπτικά μέχρι να μείνει μόνο μια συστάδα:
 - Συγχώνευση των δύο κοντινότερων συστάδων
 - Ενημέρωση του πίνακα γειτνίασης
- Βασική λειτουργία ο **υπολογισμός** της γειτνίασης δύο συστάδων

Ιεραρχική Συσταδοποίηση

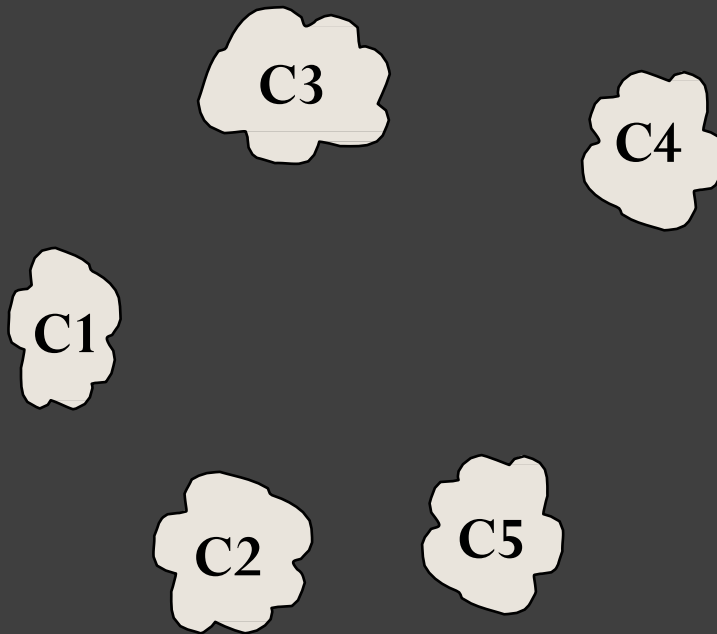
- Αρχικά κάθε σημείο αποτελεί και μία συστάδα
- Δημιουργείται ο πίνακας γειτνίασης με τις αποστάσεις των ανά δυο σημείων



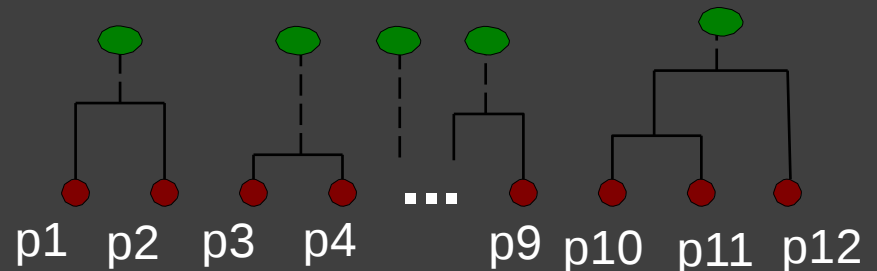
	p1	p2	p3	p4	p5	...
p1						
p2						
p3						
p4						
p5						
.						
.						
.						

Ιεραρχική Συσταδοποίηση

Μετά από κάποιες συγχωνεύσεις έχουμε 5 συστάδες



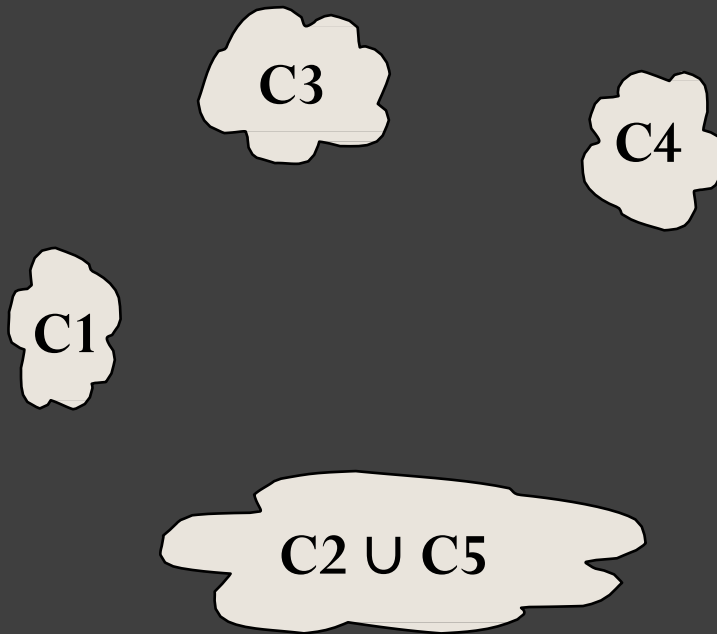
	C1	C2	C3	C4	C5
C1					
C2					
C3					
C4					
C5					



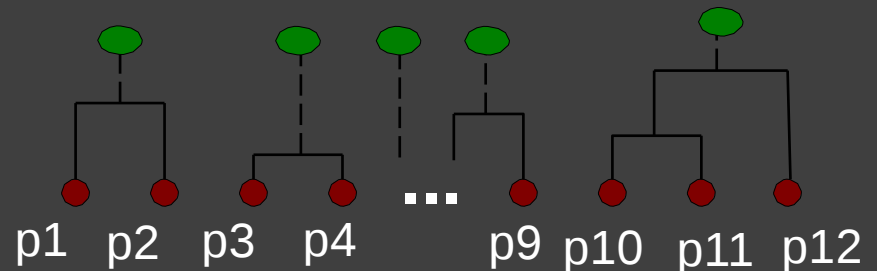
Ιεραρχική Συσταδοποίηση

Οι δύο κοντινότερες συστάδες είναι οι C2 και C5.

Τις συγχωνεύουμε και ενημερώνουμε τον πίνακα γειτνίασης



	C1	C2 U C5	C3	C4
C1				
C2 U C5				
C3				
C4				



Ιεραρχική Συσταδοποίηση

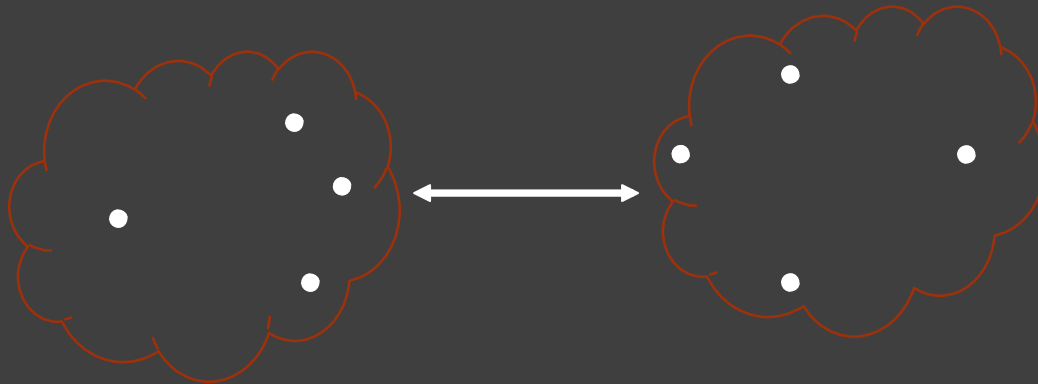
Βασικός αλγόριθμος

Έστω ότι αρχικά κάθε σημείο αποτελεί και μία συστάδα

- Υπολογισμός του πίνακα γειτνίασης
- Εκτέλεση των παρακάτω βημάτων επαναληπτικά μέχρι να μείνει μόνο μια συστάδα:
 - Συγχώνευση των δύο **κοντινότερων** συστάδων
 - **Ενημέρωση** του πίνακα γειτνίασης
- Βασική λειτουργία ο **υπολογισμός** της γειτνίασης δύο συστάδων

Υπολογισμός απόστασης

- ✓ Όταν έχω **μόνο σημεία**, επιλέγω έναν τύπο απόστασης (π.χ. Ευκλείδεια), υπολογίζω την απόσταση κάθε σημείου με όλα τα υπόλοιπα και επιλέγω τα δυο με την μικρότερη απόσταση για να γίνουν ομάδα
- ? Όταν έχουν δημιουργηθεί ομάδες, και έχω πλέον και σημεία και ομάδες ή μόνο ομάδες, προκύπτει το ερώτημα, **πώς θα υπολογίσω την απόσταση μεταξύ τους;**

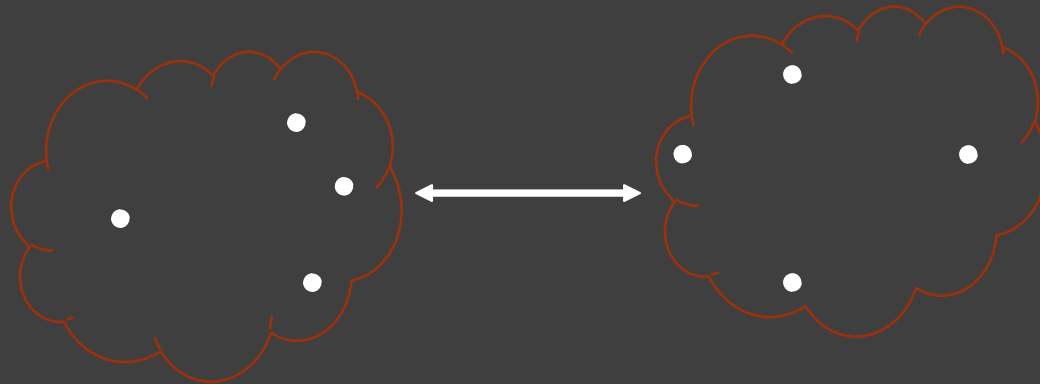


Υπολογισμός απόστασης

Άρα:

ΒΗΜΑ 1: Για να συγχωνεύσω συστάδες επιλέγω ΠΑΝΤΑ αυτές με τη **μικρότερη** απόσταση

ΒΗΜΑ 2: Για να ενημερώσω τον πίνακα απόστασης πρέπει να ορίσω ένα **κριτήριο** επιλογής της απόστασης μεταξύ των συστάδων



Υπολογισμός απόστασης

Άρα:

ΒΗΜΑ 1: Για να συγχωνεύσω συστάδες επιλέγω ΠΑΝΤΑ αυτές με τη μικρότερη απόσταση

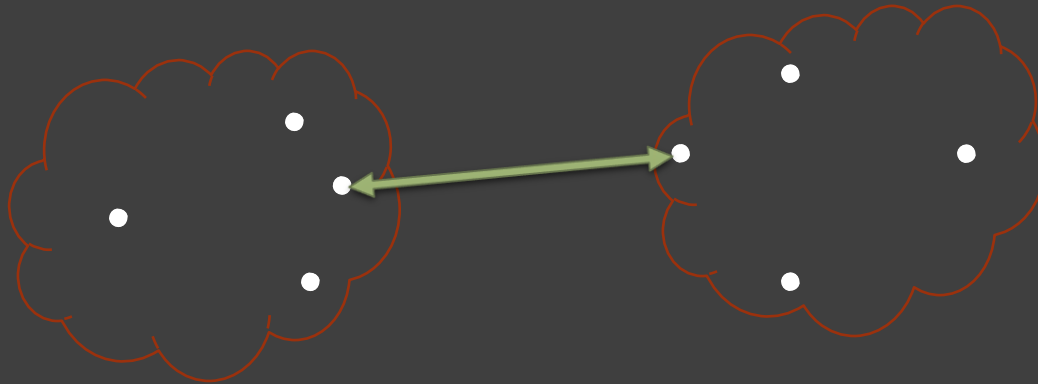
ΒΗΜΑ 2: Για να ενημερώσω τον πίνακα απόστασης **πρέπει να ορίσω** ένα **κριτήριο** επιλογής της απόστασης μεταξύ των συστάδων

- MIN
- MAX
- Μέσος όρος συστάδας
- Απόσταση μεταξύ κεντρικών σημείων

Υπολογισμός απόστασης μεταξύ συστάδων

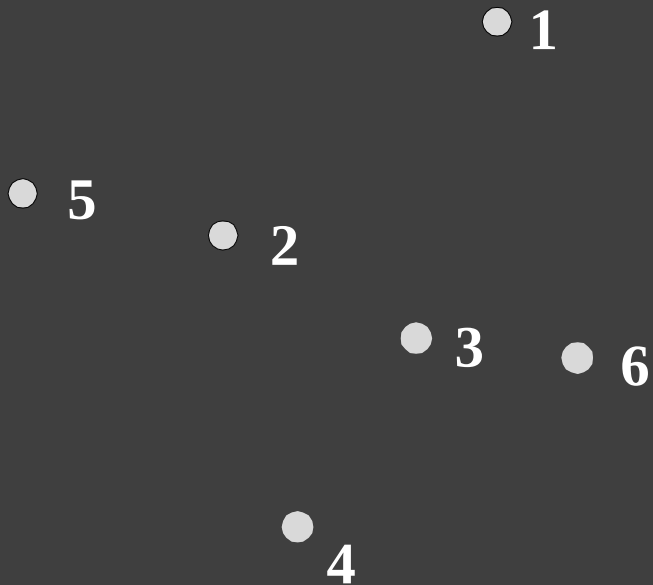
MIN

- ✓ MIN ή μοναδικής ακμής ή απλού συνδέσμου
- ✓ Η απόσταση μεταξύ δύο συστάδων ή συστάδας και σημείου ορίζεται ως η απόσταση μεταξύ των **ΚΟΝΤΙΝΌΤΕΡΩΝ** σημείων που ανήκουν στις διαφορετικές συστάδες
- ✓ Ονομάζεται και μέθοδος συσταδοποίησης κοντινότερου γείτονα



Υπολογισμός απόστασης μεταξύ συστάδων

MIN



1 (0.4, 0.53)
2 (0.22, 0.38)
3 (0.35, 0.32)
4 (0.26, 0.19)
5 (0.08, 0.41)
6 (0.45, 0.30)

Πίνακας απόστασης

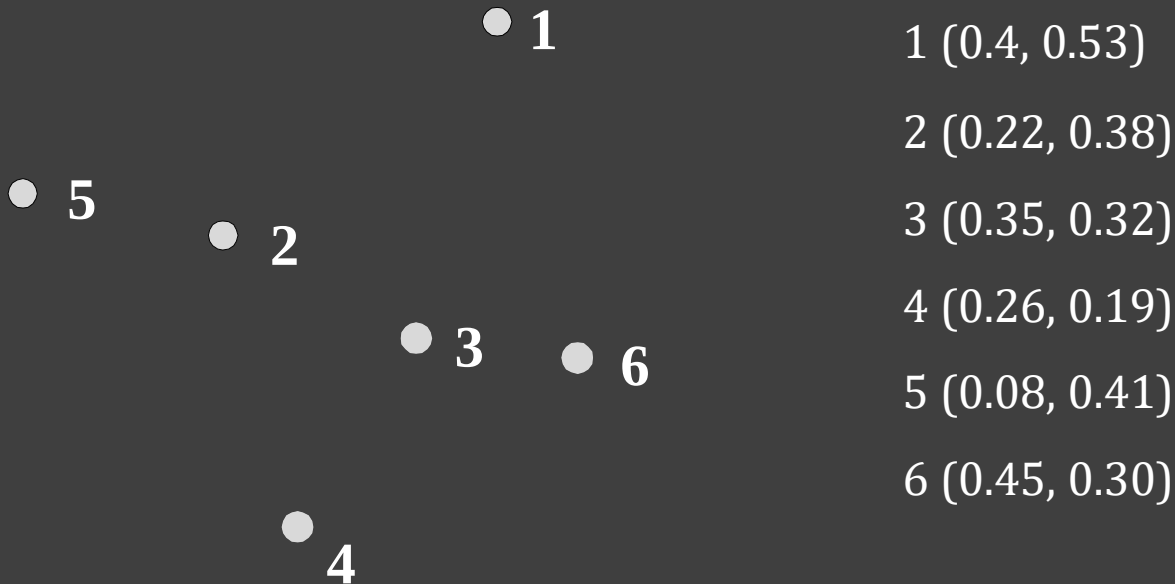
	p1	p2	p3	p4	p5	p6
p1	0.00	0.24	0.22	0.37	0.34	0.23
p2	0.24	0.00	0.15	0.20	0.14	0.25
p3	0.22	0.15	0.00	0.15	0.28	0.11
p4	0.37	0.20	0.15	0.00	0.29	0.22
p5	0.34	0.14	0.28	0.29	0.00	0.39
p6	0.23	0.25	0.11	0.22	0.39	0.00

ΒΗΜΑ 1: δεν έχω
συστάδες, μόνο σημεία
→ συγχωνεύω τις
συστάδες με τη
μικρότερη απόσταση
(ΠΑΝΤΑ)

ΒΗΜΑ 2: Για να
ενημερώσω τον πίνακα:
κριτήριο υπολογισμού
απόστασης μεταξύ των
συστάδων → **MIN**

Υπολογισμός απόστασης μεταξύ συστάδων

MIN



Πίνακας απόστασης

	p1	p2	p3	p4	p5	p6
p1	0.00	0.24	0.22	0.37	0.34	0.23
p2	0.24	0.00	0.15	0.20	0.14	0.25
p3	0.22	0.15	0.00	0.15	0.28	0.11
p4	0.37	0.20	0.15	0.00	0.29	0.22
p5	0.34	0.14	0.28	0.29	0.00	0.39
p6	0.23	0.25	0.11	0.22	0.39	0.00

min

Υπολογισμός απόστασης μεταξύ συστάδων

MIN



1 (0.4, 0.53)
2 (0.22, 0.38)
3 (0.35, 0.32)
4 (0.26, 0.19)
5 (0.08, 0.41)
6 (0.45, 0.30)

Πίνακας απόστασης

	p1	p2	p3	p4	p5	p6
p1	0.00	0.24	0.22	0.37	0.34	0.23
p2	0.24	0.00	0.15	0.20	0.14	0.25
p3	0.22	0.15	0.00	0.15	0.28	0.11
p4	0.37	0.20	0.15	0.00	0.29	0.22
p5	0.34	0.14	0.28	0.29	0.00	0.39
p6	0.23	0.25	0.11	0.22	0.39	0.00

Υπολογισμός απόστασης μεταξύ συστάδων

MIN



1 (0.4, 0.53)

2 (0.22, 0.38)

3 (0.35, 0.32)

4 (0.26, 0.19)

5 (0.08, 0.41)

6 (0.45, 0.30)

Πίνακας απόστασης

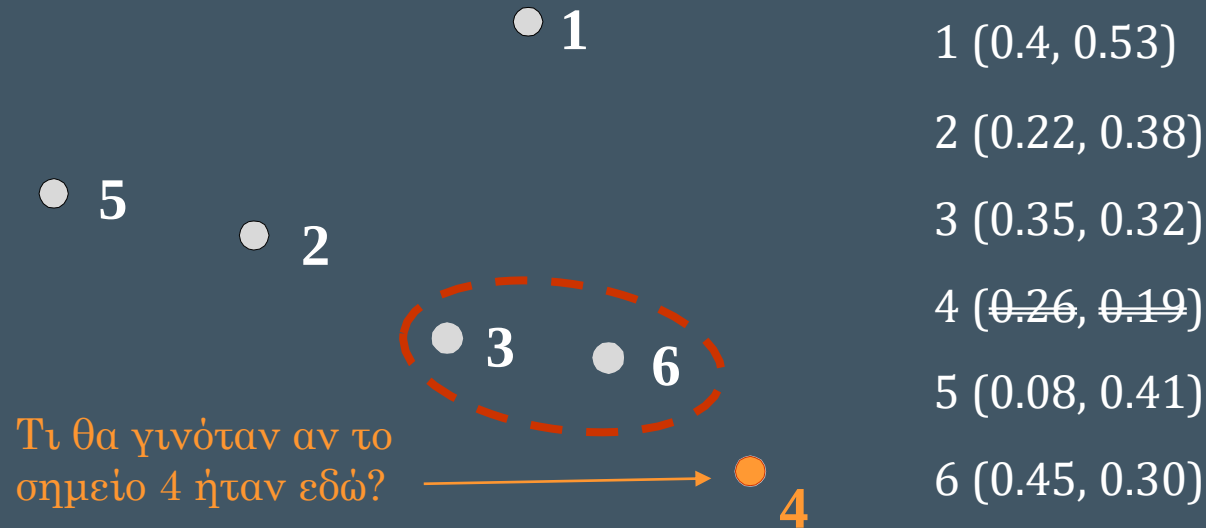
	p1	p2	p3	p4	p5	p6
p1	0.00	0.24	0.22	0.37	0.34	0.23
p2	0.24	0.00	0.15	0.20	0.14	0.25
p3	0.22	0.15	0.00	0.15	0.28	0.11
p4	0.37	0.20	0.15	0.00	0.29	0.22
p5	0.34	0.14	0.28	0.29	0.00	0.39
p6	0.23	0.25	0.11	0.22	0.39	0.00

ΒΗΜΑ 1: συγχωνεύω τις συστάδες με τη μικρότερη απόσταση (ΠΑΝΤΑ)

ΒΗΜΑ 2: για να ενημερώσω τον πίνακα: κριτήριο υπολογισμού απόστασης μεταξύ των συστάδων → **MIN**

Υπολογισμός απόστασης μεταξύ συστάδων

MIN



Πίνακας απόστασης

	p1	p2	p3	p4	p5	p6
p1	0.00	0.24	0.22	0.37	0.34	0.23
p2	0.24	0.00	0.15	0.20	0.14	0.25
p3	0.22	0.15	0.00	0.15	0.28	0.11
p4	0.37	0.20	0.15	0.00	0.29	0.22
p5	0.34	0.14	0.28	0.29	0.00	0.39
p6	0.23	0.25	0.11	0.22	0.39	0.00

ΒΗΜΑ 1: συγχωνεύω τις συστάδες με τη μικρότερη απόσταση (ΠΑΝΤΑ)

ΒΗΜΑ 2: για να ενημερώσω τον πίνακα: κριτήριο υπολογισμού απόστασης μεταξύ των συστάδων → **MIN**

Υπολογισμός απόστασης μεταξύ συστάδων

MIN



1 (0.4, 0.53)

2 (0.22, 0.38)

3 (0.35, 0.32)

4 (0.26, 0.19)

5 (0.08, 0.41)

6 (0.45, 0.30)

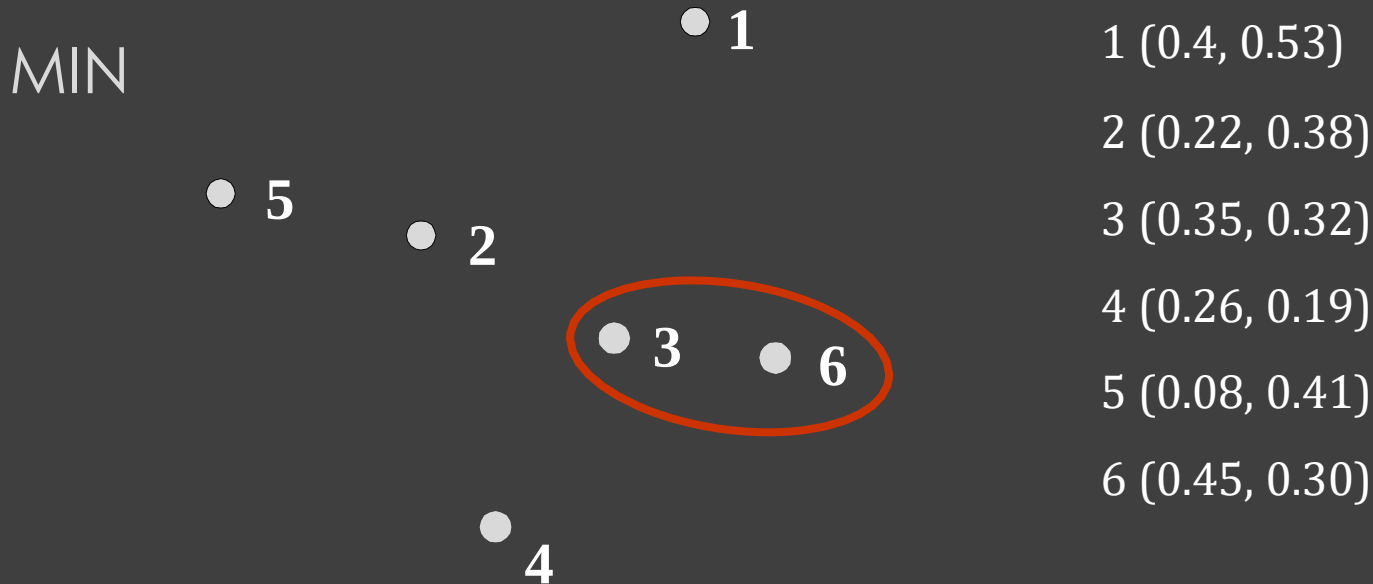
Πίνακας απόστασης

	p1	p2	p3	p4	p5	p6
p1	0.00	0.24	0.22	0.37	0.34	0.23
p2	0.24	0.00	0.15	0.20	0.14	0.25
p3	0.22	0.15	0.00	0.15	0.28	0.11
p4	0.37	0.20	0.15	0.00	0.29	0.22
p5	0.34	0.14	0.28	0.29	0.00	0.39
p6	0.23	0.25	0.11	0.22	0.39	0.00

ΒΗΜΑ 1: συγχωνεύω τις συστάδες με τη μικρότερη απόσταση (ΠΑΝΤΑ)

ΒΗΜΑ 2: για να ενημερώσω τον πίνακα: κριτήριο υπολογισμού απόστασης μεταξύ των συστάδων → **MIN**

Υπολογισμός απόστασης μεταξύ συστάδων



Πίνακας απόστασης

	p1	p2	p3 ∪ p6	p4	p5
p1	0.00	0.24	0.22	0.37	0.34
p2	0.24	0.00	0.15	0.20	0.14
p3 ∪ p6	0.22	0.15	0.00	0.15	0.28
p4	0.37	0.20	0.15	0.00	0.29
p5	0.34	0.14	0.28	0.29	0.00

ΒΗΜΑ 1: συγχωνεύω τις συστάδες με τη μικρότερη απόσταση (ΠΑΝΤΑ)

ΒΗΜΑ 2: για να ενημερώσω τον πίνακα: κριτήριο υπολογισμού απόστασης μεταξύ των συστάδων → **MIN**

Υπολογισμός απόστασης μεταξύ συστάδων

MIN



1 (0.4, 0.53)
2 (0.22, 0.38)
3 (0.35, 0.32)
4 (0.26, 0.19)
5 (0.08, 0.41)
6 (0.45, 0.30)

Πίνακας απόστασης

	p1	p2	p3 ∪ p6	p4	p5
p1	0.00	0.24	0.22	0.37	0.34
p2	0.24	0.00	0.15	0.20	0.14
p3 ∪ p6	0.22	0.15	0.00	0.15	0.28
p4	0.37	0.20	0.15	0.00	0.29
p5	0.34	0.14	0.28	0.29	0.00

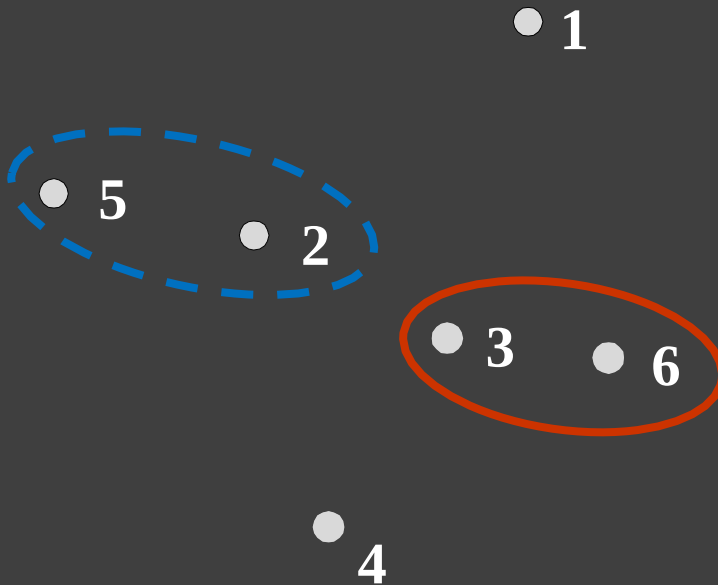
min

ΒΗΜΑ 1: συγχωνεύω τις συστάδες με τη μικρότερη απόσταση (ΠΑΝΤΑ)

ΒΗΜΑ 2: Για να ενημερώσω τον πίνακα: κριτήριο υπολογισμού απόστασης μεταξύ των συστάδων → **MIN**

Υπολογισμός απόστασης μεταξύ συστάδων

MIN



1 (0.4, 0.53)
2 (0.22, 0.38)
3 (0.35, 0.32)
4 (0.26, 0.19)
5 (0.08, 0.41)
6 (0.45, 0.30)

Πίνακας απόστασης

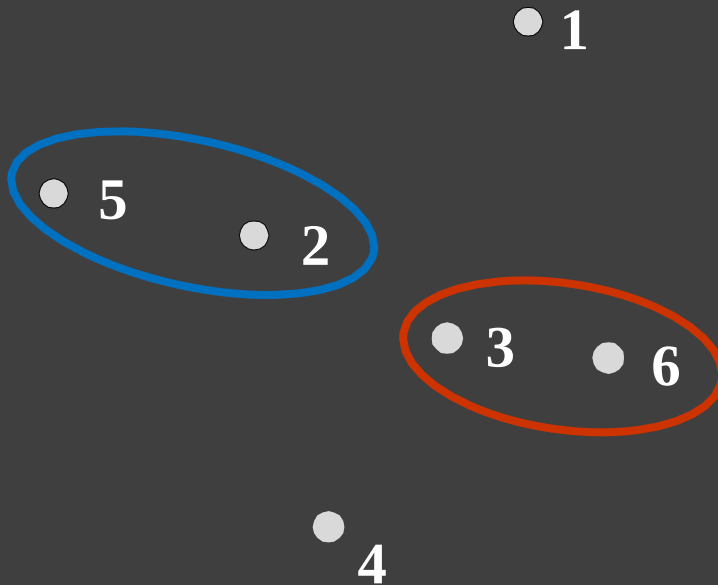
	p1	p2	p3 ∪ p6	p4	p5
p1	0.00	0.24	0.22	0.37	0.34
p2	0.24	0.00	0.15	0.20	0.14
p3 ∪ p6	0.22	0.15	0.00	0.15	0.28
p4	0.37	0.20	0.15	0.00	0.29
p5	0.34	0.14	0.28	0.29	0.00

ΒΗΜΑ 1: συγχωνεύω τις συστάδες με τη μικρότερη απόσταση (ΠΑΝΤΑ)

ΒΗΜΑ 2: για να ενημερώσω τον πίνακα: κριτήριο υπολογισμού απόστασης μεταξύ των συστάδων → **MIN**

Υπολογισμός απόστασης μεταξύ συστάδων

MIN



1 (0.4, 0.53)
2 (0.22, 0.38)
3 (0.35, 0.32)
4 (0.26, 0.19)
5 (0.08, 0.41)
6 (0.45, 0.30)

Πίνακας απόστασης

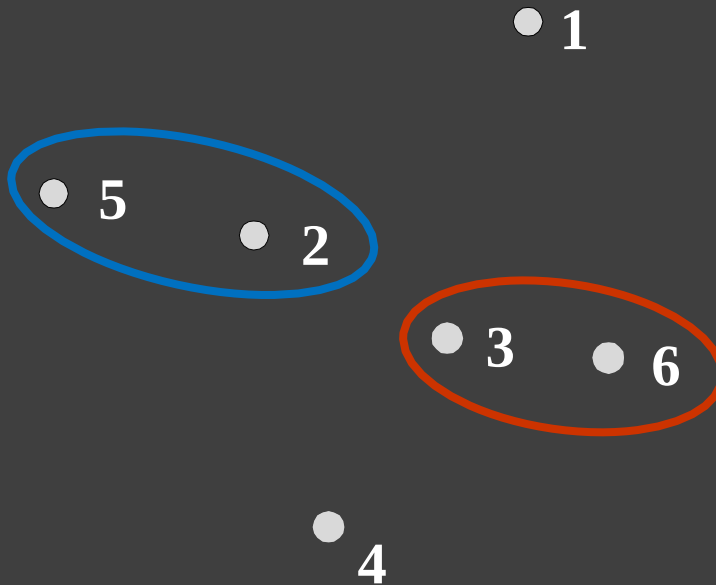
	p1	p2 $\cup$ p5	p3 $\cup$ p6	p4
p1	0.00	0.24	0.22	0.37
p2 $\cup$ p5	0.24	0.00	0.15	0.20
p3 $\cup$ p6	0.22	0.15	0.00	0.15
p4	0.37	0.20	0.15	0.00

ΒΗΜΑ 1: συγχωνεύω τις συστάδες με τη μικρότερη απόσταση (ΠΑΝΤΑ)

ΒΗΜΑ 2: για να ενημερώσω τον πίνακα: κριτήριο υπολογισμού απόστασης μεταξύ των συστάδων → **MIN**

Υπολογισμός απόστασης μεταξύ συστάδων

MIN



1 (0.4, 0.53)
2 (0.22, 0.38)
3 (0.35, 0.32)
4 (0.26, 0.19)
5 (0.08, 0.41)
6 (0.45, 0.30)

Πίνακας απόστασης

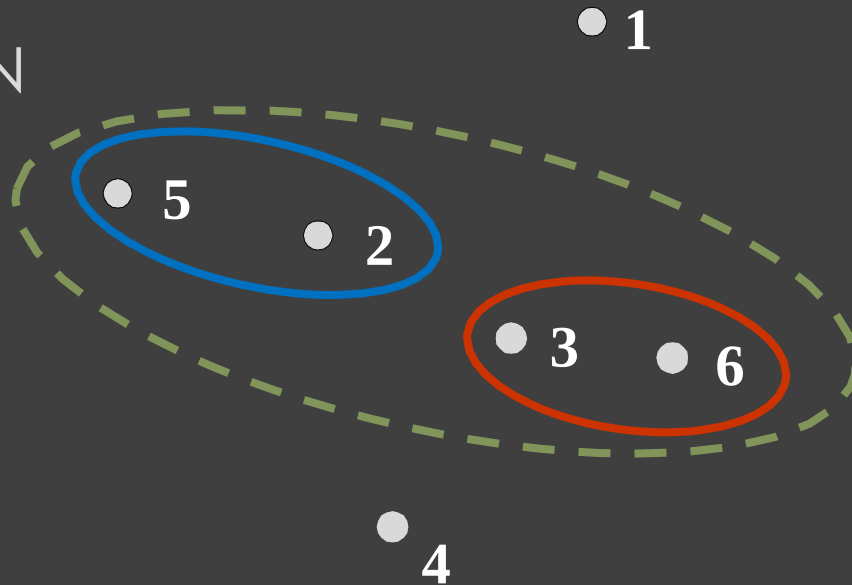
	p1	p2 ∪ p5	p3 ∪ p6	p4
p1	0.00	0.24	0.22	0.37
p2 ∪ p5	0.24	0.00	0.15	0.20
p3 ∪ p6	0.22	0.15	0.00	0.15
p4	0.37	0.20	0.15	0.00

ΒΗΜΑ 1: συγχωνεύω τις συστάδες με τη μικρότερη απόσταση (ΠΑΝΤΑ)

ΒΗΜΑ 2: Για να ενημερώσω τον πίνακα: κριτήριο υπολογισμού απόστασης μεταξύ των συστάδων → **MIN**

Υπολογισμός απόστασης μεταξύ συστάδων

MIN



1 (0.4, 0.53)
2 (0.22, 0.38)
3 (0.35, 0.32)
4 (0.26, 0.19)
5 (0.08, 0.41)
6 (0.45, 0.30)

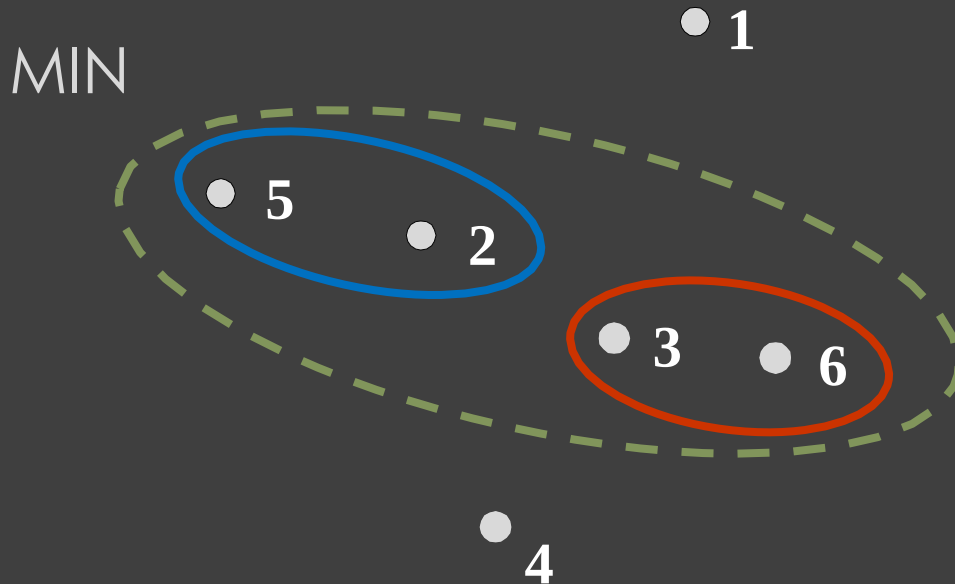
Πίνακας απόστασης

	p1	p2 ∪ p5	p3 ∪ p6	p4
p1	0.00	0.24	0.22	0.37
p2 ∪ p5	0.24	0.00	0.15	0.20
p3 ∪ p6	0.22	0.15	0.00	0.15
p4	0.37	0.20	0.15	0.00

Ισότητα

Επιλέγουμε τυχαία ένα από τα δυο
έστω αυτό

Υπολογισμός απόστασης μεταξύ συστάδων



1 (0.4, 0.53)
2 (0.22, 0.38)
3 (0.35, 0.32)
4 (0.26, 0.19)
5 (0.08, 0.41)
6 (0.45, 0.30)

Πίνακας απόστασης

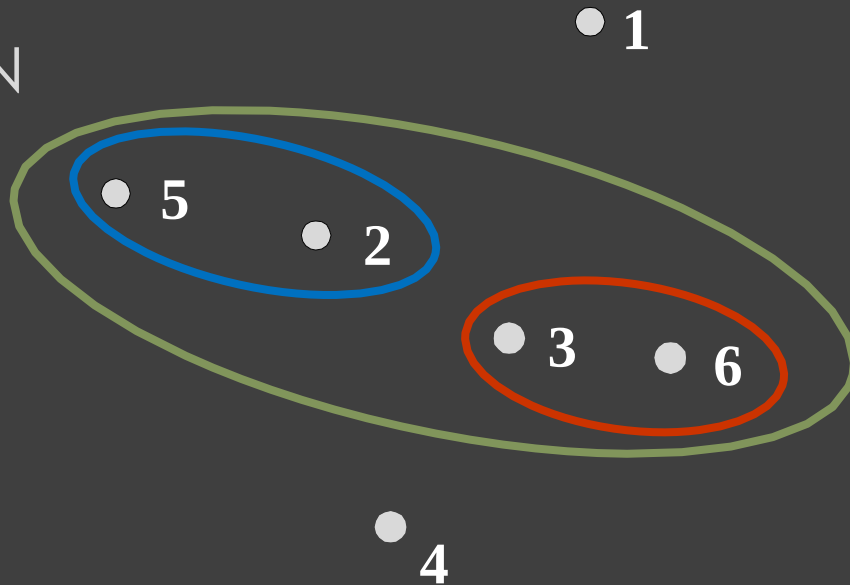
	p1	p2 ∪ p5	p3 ∪ p6	p4
p1	0.00	0.24	0.22	0.37
p2 ∪ p5	0.24	0.00	0.15	0.20
p3 ∪ p6	0.22	0.15	0.00	0.15
p4	0.37	0.20	0.15	0.00

ΒΗΜΑ 1: συγχωνεύω τις συστάδες με τη μικρότερη απόσταση (ΠΑΝΤΑ)

ΒΗΜΑ 2: για να ενημερώσω τον πίνακα: κριτήριο υπολογισμού απόστασης μεταξύ των συστάδων → **MIN**

Υπολογισμός απόστασης μεταξύ συστάδων

MIN



1 (0.4, 0.53)

2 (0.22, 0.38)

3 (0.35, 0.32)

4 (0.26, 0.19)

5 (0.08, 0.41)

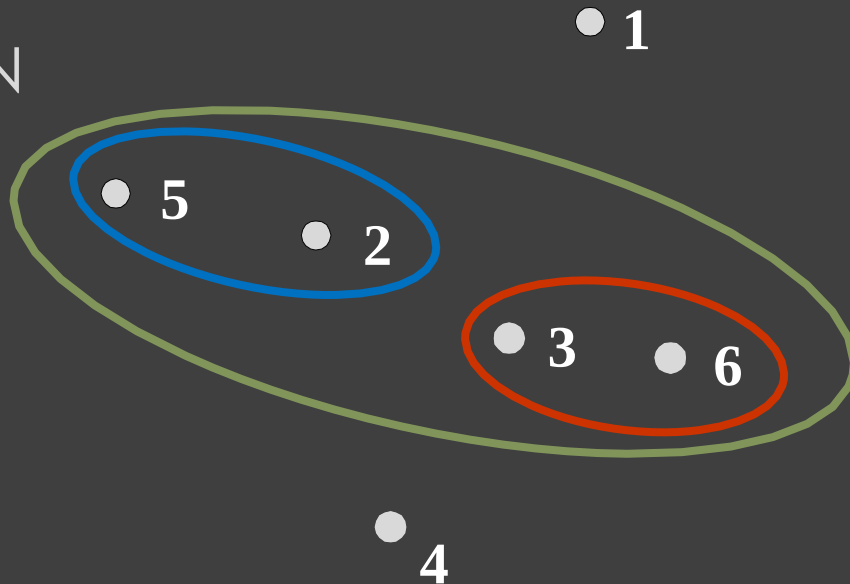
6 (0.45, 0.30)

Πίνακας απόστασης

	p1	p2 $\cup$ p5 $\cup$ p3 $\cup$ p6	p4
p1	0.00	0.22	0.37
p2 $\cup$ p5 $\cup$ p3 $\cup$ p6	0.22	0.00	0.15
p4	0.37	0.15	0.00

Υπολογισμός απόστασης μεταξύ συστάδων

MIN



1 (0.4, 0.53)

2 (0.22, 0.38)

3 (0.35, 0.32)

4 (0.26, 0.19)

5 (0.08, 0.41)

6 (0.45, 0.30)

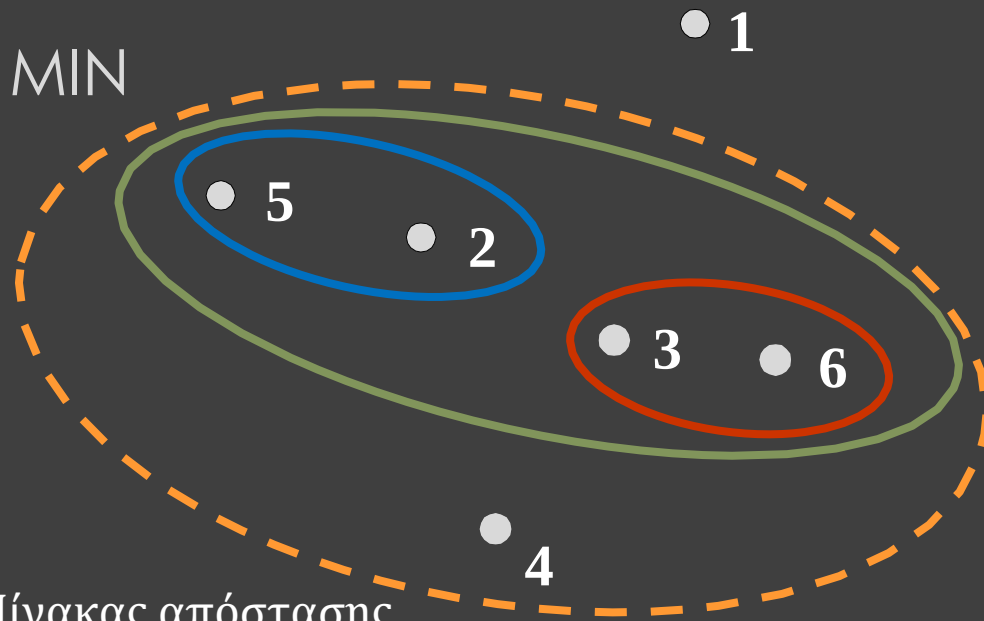
Πίνακας απόστασης

	p1	p2 $\cup$ p5 $\cup$ p3 $\cup$ p6	p4
p1	0.00	0.22	0.37
p2 $\cup$ p5 $\cup$ p3 $\cup$ p6	0.22	0.00	0.15
p4	0.37	0.15	0.00

ΒΗΜΑ 1: συγχωνεύω τις συστάδες με τη μικρότερη απόσταση (ΠΑΝΤΑ)

ΒΗΜΑ 2: Για να ενημερώσω τον πίνακα: κριτήριο υπολογισμού απόστασης μεταξύ των συστάδων $\rightarrow$ **MIN**

Υπολογισμός απόστασης μεταξύ συστάδων



1 (0.4, 0.53)
2 (0.22, 0.38)
3 (0.35, 0.32)
4 (0.26, 0.19)
5 (0.08, 0.41)
6 (0.45, 0.30)

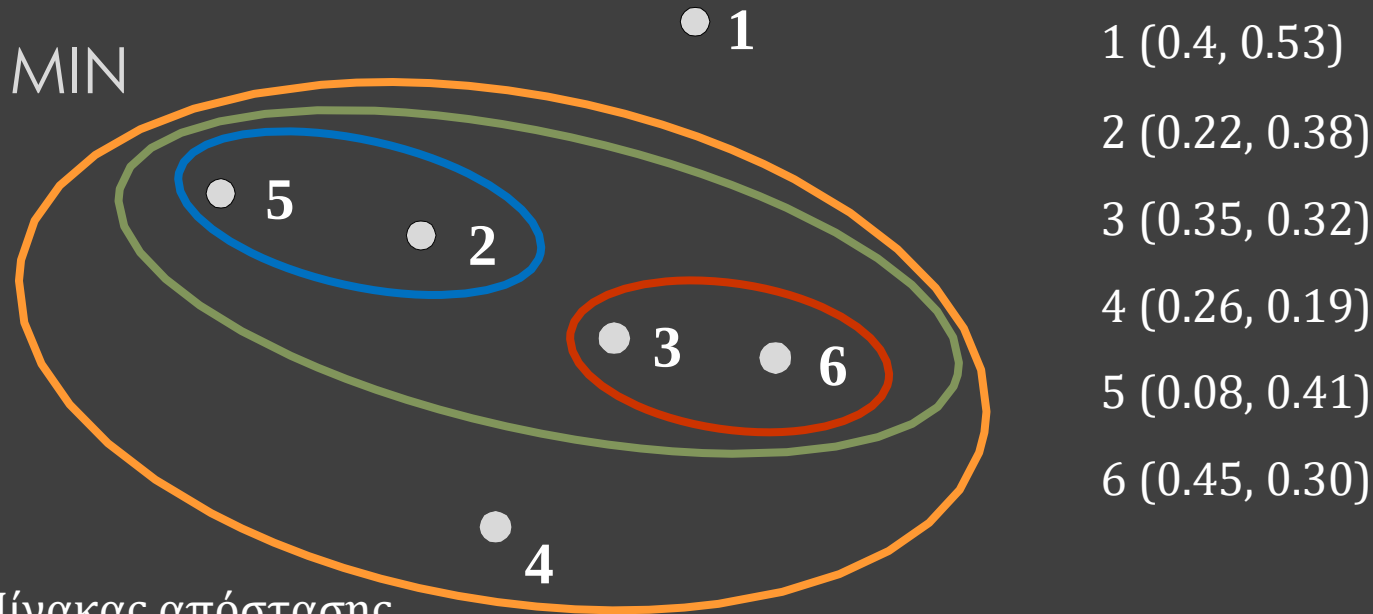
Πίνακας απόστασης

	p1	p2 ∪ p5 ∪ p3 ∪ p6	p4
p1	0.00	0.22	0.37
p2 ∪ p5 ∪ p3 ∪ p6	0.22	0.00	0.15
p4	0.37	0.15	0.00

ΒΗΜΑ 1: συγχωνεύω τις συστάδες με τη μικρότερη απόσταση (ΠΑΝΤΑ)

ΒΗΜΑ 2: για να ενημερώσω τον πίνακα: κριτήριο υπολογισμού απόστασης μεταξύ των συστάδων → **MIN**

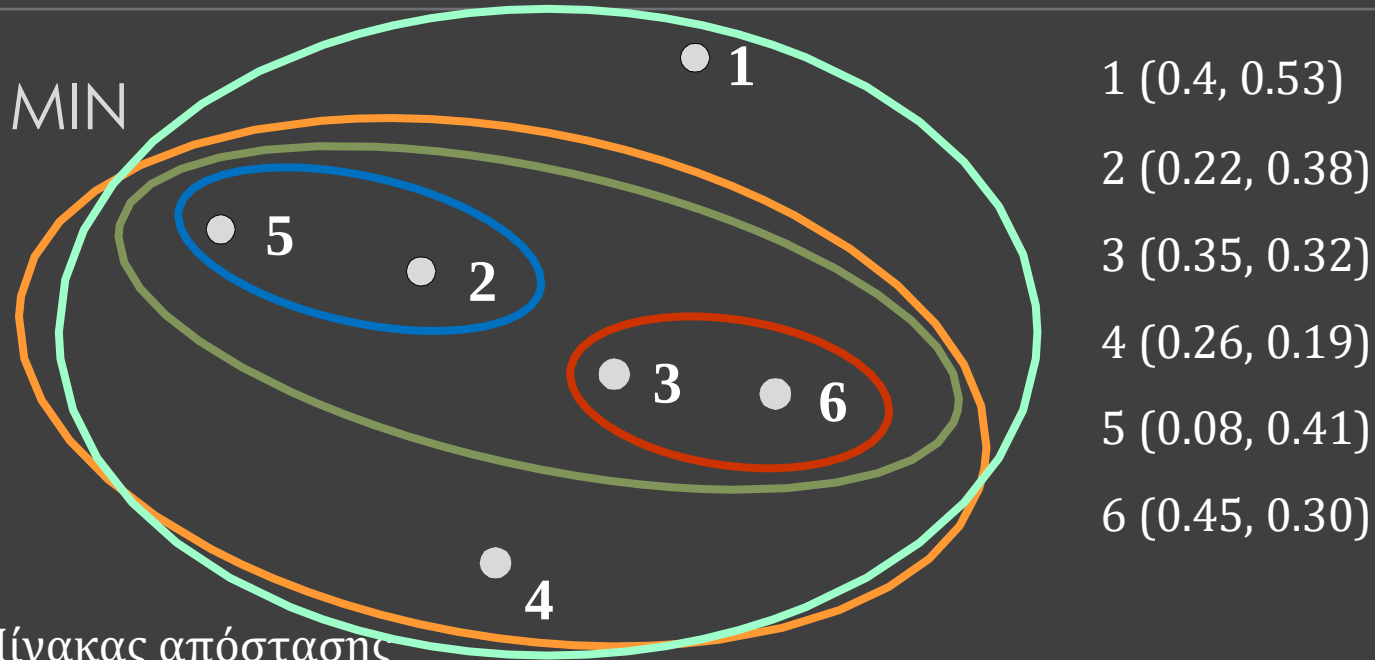
Υπολογισμός απόστασης μεταξύ συστάδων



Πίνακας απόστασης

	p1	p2 ∪ p5 ∪ p3 ∪ p6 ∪ p4
p1	0.00	0.22
p2 ∪ p5 ∪ p3 ∪ p6 ∪ p4	0.22	0.00

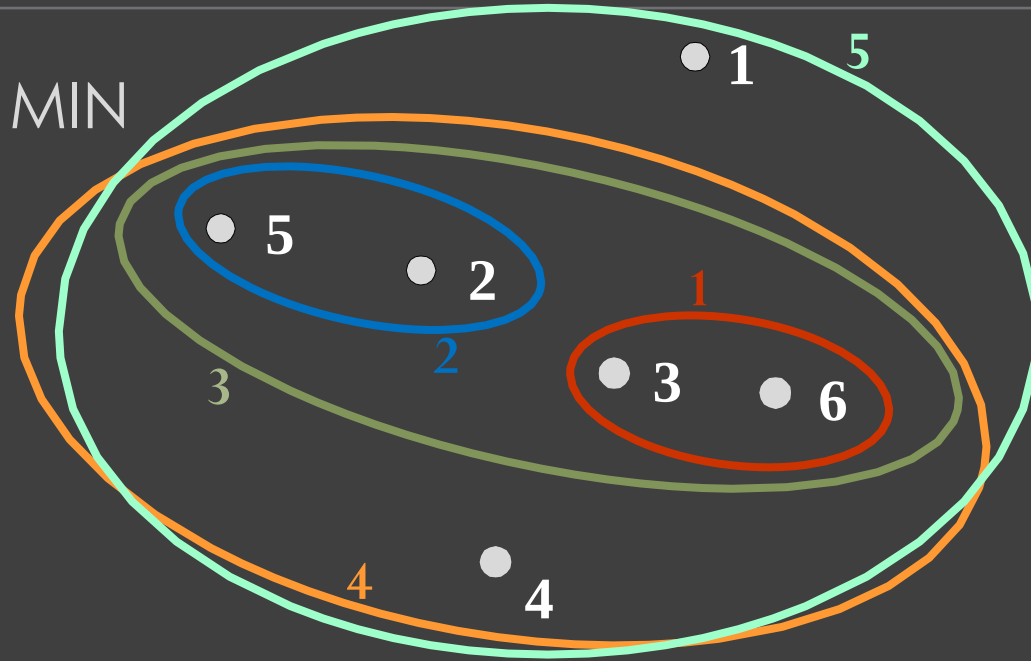
Υπολογισμός απόστασης μεταξύ συστάδων



Πίνακας απόστασης

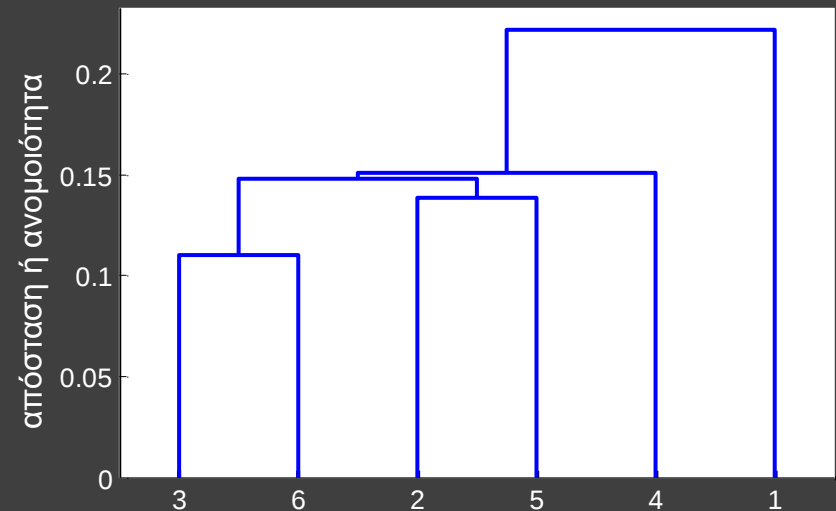
	$p1 \cup p2 \cup p5 \cup p3 \cup p6 \cup p4$
$p1 \cup p2 \cup p5 \cup p3 \cup p6 \cup p4$	0.00

Υπολογισμός απόστασης μεταξύ συστάδων



Σειρά συγχωνεύσεων:

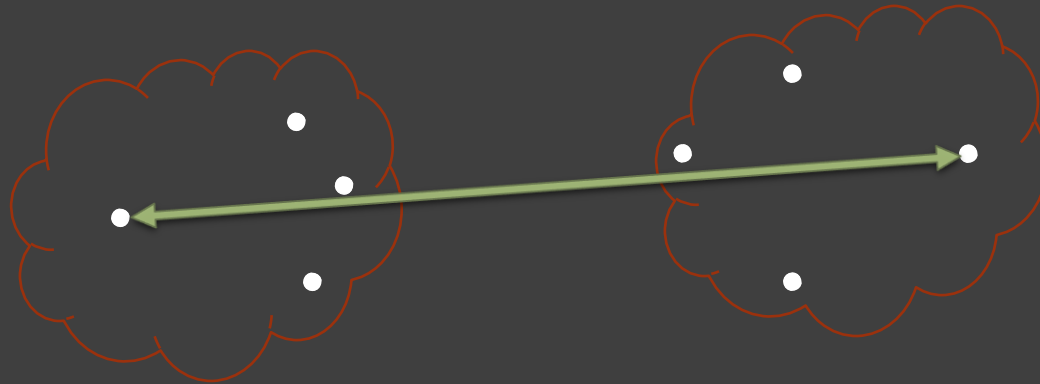
1. το σημείο 3 με το σημείο 6 (0.11)
2. το σημείο 2 με το σημείο 5 (0.14)
3. η ομάδα 3-6 με την ομάδα 2-5 (0.15)
4. η ομάδα 2-3-5-6 με το σημείο 4 (0.15)
5. η ομάδα 2-3-4-5-6 με το σημείο 1 (0.22)



Υπολογισμός απόστασης μεταξύ συστάδων

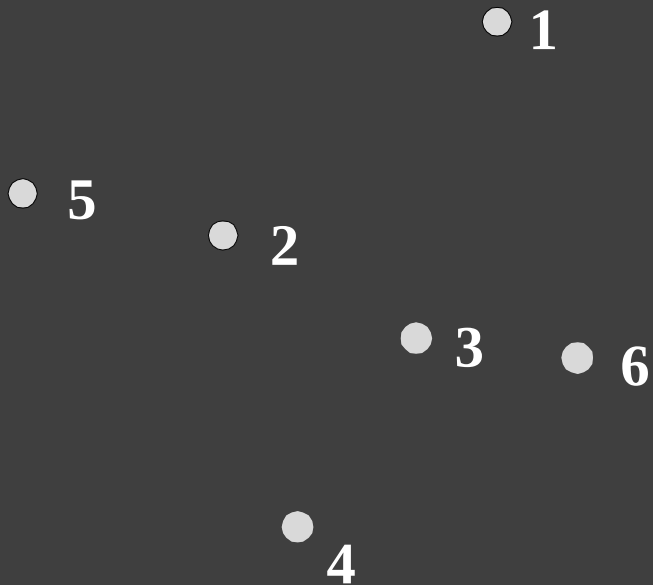
MAX

- ✓ Η απόσταση μεταξύ δύο συστάδων ή συστάδας και σημείου ορίζεται ως η απόσταση μεταξύ των **πιο απομακρυσμένων** σημείων που ανήκουν στις διαφορετικές συστάδες



Υπολογισμός απόστασης μεταξύ συστάδων

MAX



1 (0.4, 0.53)
2 (0.22, 0.38)
3 (0.35, 0.32)
4 (0.26, 0.19)
5 (0.08, 0.41)
6 (0.45, 0.30)

Πίνακας απόστασης

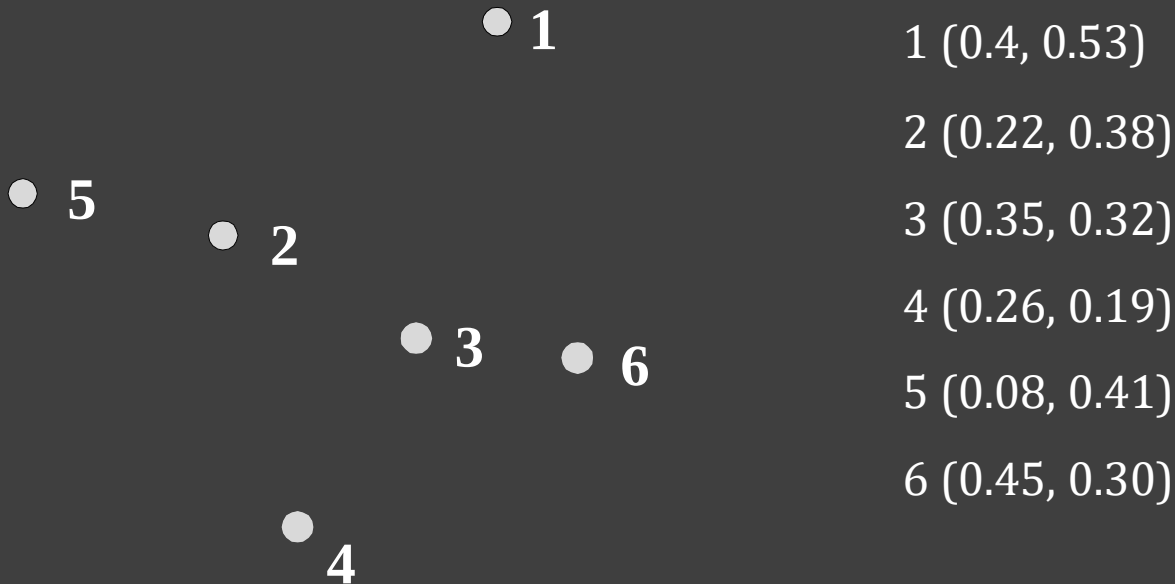
	p1	p2	p3	p4	p5	p6
p1	0.00	0.24	0.22	0.37	0.34	0.23
p2	0.24	0.00	0.15	0.20	0.14	0.25
p3	0.22	0.15	0.00	0.15	0.28	0.11
p4	0.37	0.20	0.15	0.00	0.29	0.22
p5	0.34	0.14	0.28	0.29	0.00	0.39
p6	0.23	0.25	0.11	0.22	0.39	0.00

ΒΗΜΑ 1: δεν έχω
συστάδες, μόνο σημεία
→ συγχωνεύω τις
συστάδες με τη
μικρότερη απόσταση
(ΠΑΝΤΑ)

ΒΗΜΑ 2: Για να
ενημερώσω τον πίνακα:
κριτήριο υπολογισμού
απόστασης μεταξύ των
συστάδων → **MAX**

Υπολογισμός απόστασης μεταξύ συστάδων

MAX



Πίνακας απόστασης

	p1	p2	p3	p4	p5	p6
p1	0.00	0.24	0.22	0.37	0.34	0.23
p2	0.24	0.00	0.15	0.20	0.14	0.25
p3	0.22	0.15	0.00	0.15	0.28	0.11
p4	0.37	0.20	0.15	0.00	0.29	0.22
p5	0.34	0.14	0.28	0.29	0.00	0.39
p6	0.23	0.25	0.11	0.22	0.39	0.00

min

Υπολογισμός απόστασης μεταξύ συστάδων

MAX



1 (0.4, 0.53)
2 (0.22, 0.38)
3 (0.35, 0.32)
4 (0.26, 0.19)
5 (0.08, 0.41)
6 (0.45, 0.30)

Πίνακας απόστασης

	p1	p2	p3	p4	p5	p6
p1	0.00	0.24	0.22	0.37	0.34	0.23
p2	0.24	0.00	0.15	0.20	0.14	0.25
p3	0.22	0.15	0.00	0.15	0.28	0.11
p4	0.37	0.20	0.15	0.00	0.29	0.22
p5	0.34	0.14	0.28	0.29	0.00	0.39
p6	0.23	0.25	0.11	0.22	0.39	0.00

Υπολογισμός απόστασης μεταξύ συστάδων

MAX



1 (0.4, 0.53)
2 (0.22, 0.38)
3 (0.35, 0.32)
4 (0.26, 0.19)
5 (0.08, 0.41)
6 (0.45, 0.30)

Πίνακας απόστασης

	p1	p2	p3	p4	p5	p6
p1	0.00	0.24	0.22	0.37	0.34	0.23
p2	0.24	0.00	0.15	0.20	0.14	0.25
p3	0.22	0.15	0.00	0.15	0.28	0.11
p4	0.37	0.20	0.15	0.00	0.29	0.22
p5	0.34	0.14	0.28	0.29	0.00	0.39
p6	0.23	0.25	0.11	0.22	0.39	0.00

ΒΗΜΑ 1: συγχωνεύω τις συστάδες με τη μικρότερη απόσταση (ΠΑΝΤΑ)

ΒΗΜΑ 2: για να ενημερώσω τον πίνακα: κριτήριο υπολογισμού απόστασης μεταξύ των συστάδων → **MAX**

Υπολογισμός απόστασης μεταξύ συστάδων

MAX



1 (0.4, 0.53)
2 (0.22, 0.38)
3 (0.35, 0.32)
4 (0.26, 0.19)
5 (0.08, 0.41)
6 (0.45, 0.30)

Πίνακας απόστασης

	p1	p2	p3 ∪ p6	p4	p5
p1	0.00	0.24	0.23	0.37	0.34
p2	0.24	0.00	0.25	0.20	0.14
p3 ∪ p6	0.23	0.25	0.00	0.22	0.39
p4	0.37	0.20	0.22	0.00	0.29
p5	0.34	0.14	0.39	0.29	0.00

ΒΗΜΑ 1: συγχωνεύω τις συστάδες με τη μικρότερη απόσταση (ΠΑΝΤΑ)

ΒΗΜΑ 2: για να ενημερώσω τον πίνακα: κριτήριο υπολογισμού απόστασης μεταξύ των συστάδων → **MAX**

Υπολογισμός απόστασης μεταξύ συστάδων

MAX



1 (0.4, 0.53)
2 (0.22, 0.38)
3 (0.35, 0.32)
4 (0.26, 0.19)
5 (0.08, 0.41)
6 (0.45, 0.30)

Πίνακας απόστασης

	p1	p2	p3 ∪ p6	p4	p5
p1	0.00	0.24	0.23	0.37	0.34
p2	0.24	0.00	0.25	0.20	0.14
p3 ∪ p6	0.23	0.25	0.00	0.22	0.39
p4	0.37	0.20	0.22	0.00	0.29
p5	0.34	0.14	0.39	0.29	0.00

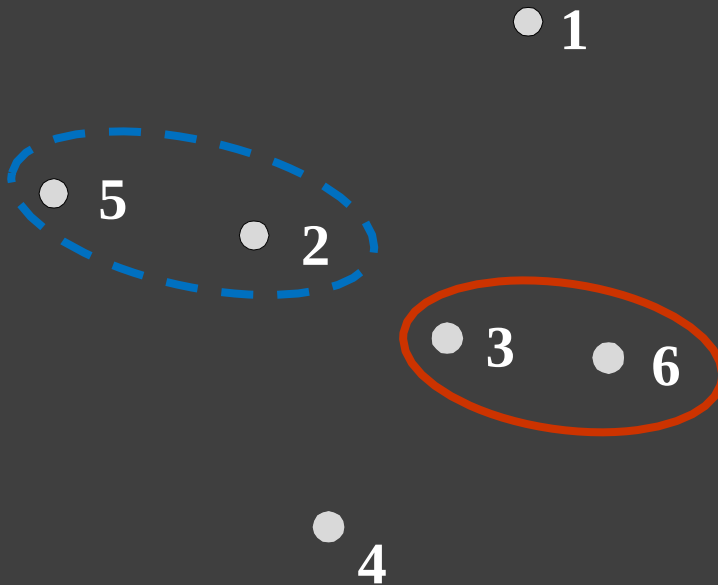
min

ΒΗΜΑ 1: συγχωνεύω τις συστάδες με τη μικρότερη απόσταση (ΠΑΝΤΑ)

ΒΗΜΑ 2: Για να ενημερώσω τον πίνακα: κριτήριο υπολογισμού απόστασης μεταξύ των συστάδων → **MAX**

Υπολογισμός απόστασης μεταξύ συστάδων

MAX



1 (0.4, 0.53)
2 (0.22, 0.38)
3 (0.35, 0.32)
4 (0.26, 0.19)
5 (0.08, 0.41)
6 (0.45, 0.30)

Πίνακας απόστασης

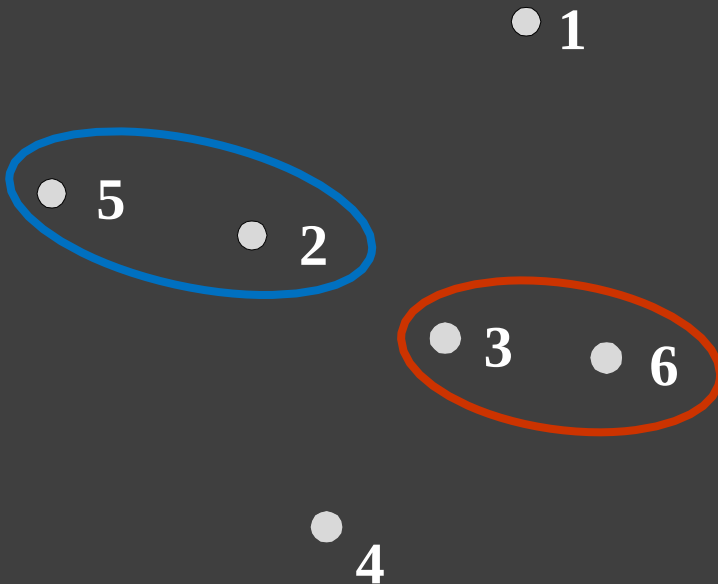
	p1	p2	p3 ∪ p6	p4	p5
p1	0.00	0.24	0.23	0.37	0.34
p2	0.24	0.00	0.25	0.20	0.14
p3 ∪ p6	0.23	0.25	0.00	0.22	0.39
p4	0.37	0.20	0.22	0.00	0.29
p5	0.34	0.14	0.39	0.29	0.00

ΒΗΜΑ 1: συγχωνεύω τις συστάδες με τη μικρότερη απόσταση (ΠΑΝΤΑ)

ΒΗΜΑ 2: για να ενημερώσω τον πίνακα: κριτήριο υπολογισμού απόστασης μεταξύ των συστάδων → **MAX**

Υπολογισμός απόστασης μεταξύ συστάδων

MAX



1 (0.4, 0.53)
2 (0.22, 0.38)
3 (0.35, 0.32)
4 (0.26, 0.19)
5 (0.08, 0.41)
6 (0.45, 0.30)

Πίνακας απόστασης

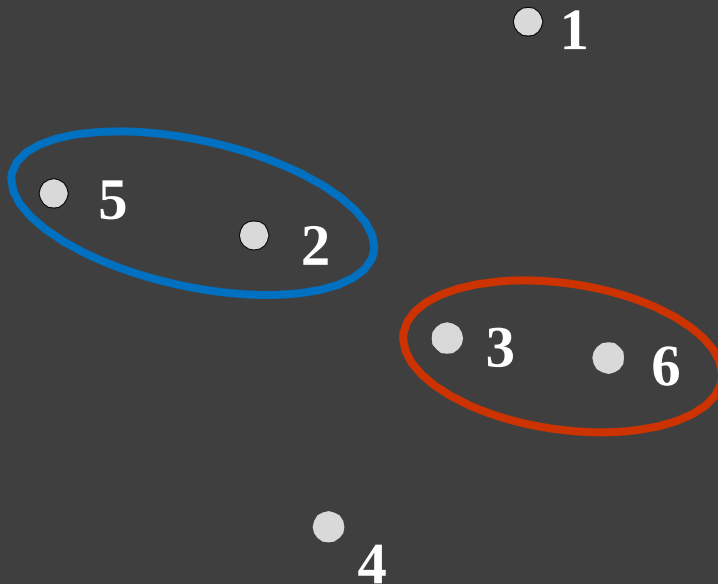
	p1	p2 $\cup$ p5	p3 $\cup$ p6	p4
p1	0.00	0.34	0.23	0.37
p2 $\cup$ p5	0.34	0.00	0.39	0.29
p3 $\cup$ p6	0.23	0.39	0.00	0.22
p4	0.37	0.29	0.22	0.00

ΒΗΜΑ 1: συγχωνεύω τις συστάδες με τη μικρότερη απόσταση (ΠΑΝΤΑ)

ΒΗΜΑ 2: για να ενημερώσω τον πίνακα: κριτήριο υπολογισμού απόστασης μεταξύ των συστάδων $\rightarrow$ **MAX**

Υπολογισμός απόστασης μεταξύ συστάδων

MAX



1 (0.4, 0.53)
2 (0.22, 0.38)
3 (0.35, 0.32)
4 (0.26, 0.19)
5 (0.08, 0.41)
6 (0.45, 0.30)

Πίνακας απόστασης

	p1	p2 ∪ p5	p3 ∪ p6	p4
p1	0.00	0.34	0.23	0.37
p2 ∪ p5	0.34	0.00	0.39	0.29
p3 ∪ p6	0.23	0.39	0.00	0.22
p4	0.37	0.29	0.22	0.00

ΒΗΜΑ 1: συγχωνεύω τις συστάδες με τη μικρότερη απόσταση (ΠΑΝΤΑ)

ΒΗΜΑ 2: Για να ενημερώσω τον πίνακα: κριτήριο υπολογισμού απόστασης μεταξύ των συστάδων → **MAX**

Υπολογισμός απόστασης μεταξύ συστάδων

MAX



1 (0.4, 0.53)
2 (0.22, 0.38)
3 (0.35, 0.32)
4 (0.26, 0.19)
5 (0.08, 0.41)
6 (0.45, 0.30)

Πίνακας απόστασης

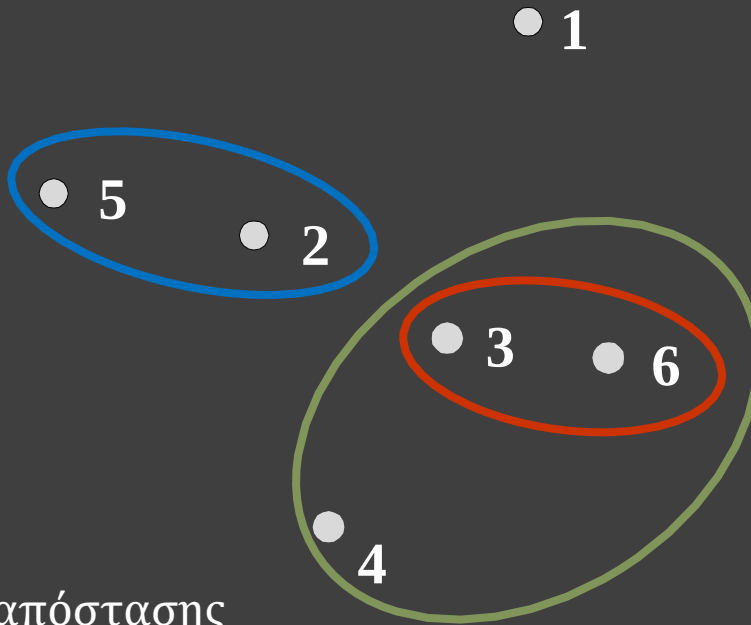
	p1	p2 $\cup$ p5	p3 $\cup$ p6	p4
p1	0.00	0.34	0.23	0.37
p2 $\cup$ p5	0.34	0.00	0.39	0.29
p3 $\cup$ p6	0.23	0.39	0.00	0.22
p4	0.37	0.29	0.22	0.00

ΒΗΜΑ 1: συγχωνεύω τις συστάδες με τη μικρότερη απόσταση (ΠΑΝΤΑ)

ΒΗΜΑ 2: για να ενημερώσω τον πίνακα: κριτήριο υπολογισμού απόστασης μεταξύ των συστάδων → **MAX**

Υπολογισμός απόστασης μεταξύ συστάδων

MAX



1 (0.4, 0.53)
2 (0.22, 0.38)
3 (0.35, 0.32)
4 (0.26, 0.19)
5 (0.08, 0.41)
6 (0.45, 0.30)

Πίνακας απόστασης

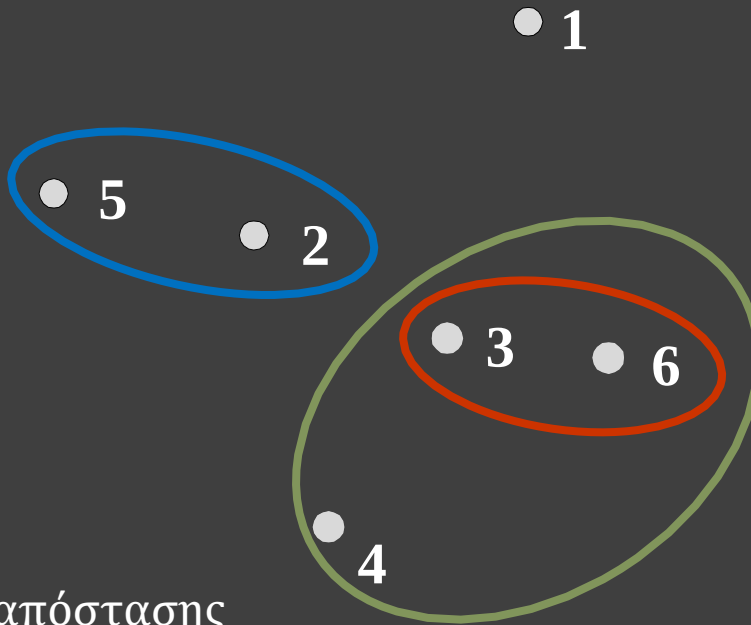
	p1	p2 $\cup$ p5	p3 $\cup$ p6 $\cup$ p4
p1	0.00	0.34	0.37
p2 $\cup$ p5	0.34	0.00	0.39
p3 $\cup$ p6 $\cup$ p4	0.37	0.39	0.00

ΒΗΜΑ 1: συγχωνεύω τις συστάδες με τη μικρότερη απόσταση (ΠΑΝΤΑ)

ΒΗΜΑ 2: Για να ενημερώσω τον πίνακα: κριτήριο υπολογισμού απόστασης μεταξύ των συστάδων $\rightarrow$ **MAX**

Υπολογισμός απόστασης μεταξύ συστάδων

MAX



1 (0.4, 0.53)

2 (0.22, 0.38)

3 (0.35, 0.32)

4 (0.26, 0.19)

5 (0.08, 0.41)

6 (0.45, 0.30)

Πίνακας απόστασης

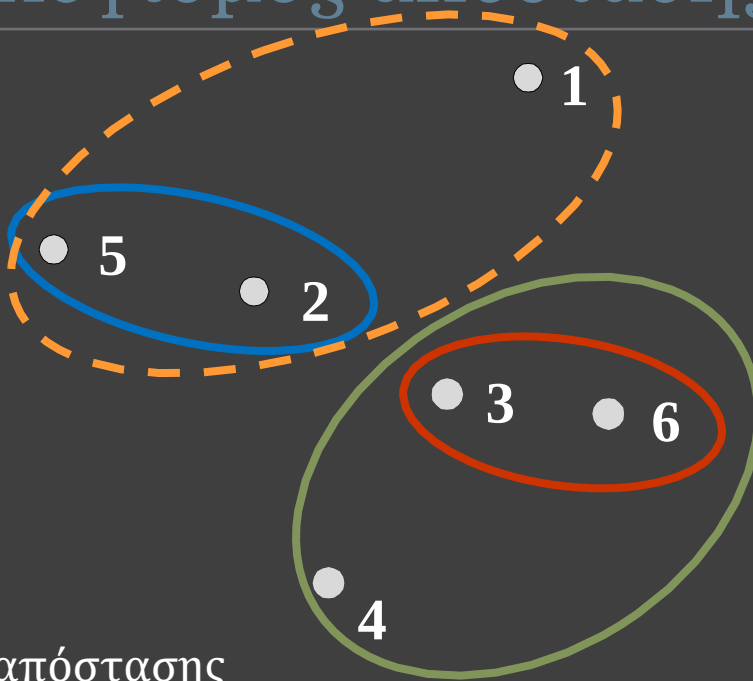
	p1	p2 $\cup$ p5	p3 $\cup$ p6 $\cup$ p4
p1	0.00	0.34	0.37
p2 $\cup$ p5	0.34	0.00	0.39
p3 $\cup$ p6 $\cup$ p4	0.37	0.39	0.00

ΒΗΜΑ 1: συγχωνεύω τις συστάδες με τη μικρότερη απόσταση (ΠΑΝΤΑ)

ΒΗΜΑ 2: Για να ενημερώσω τον πίνακα: κριτήριο υπολογισμού απόστασης μεταξύ των συστάδων $\rightarrow$ **MAX**

Υπολογισμός απόστασης μεταξύ συστάδων

MAX



1 (0.4, 0.53)
2 (0.22, 0.38)
3 (0.35, 0.32)
4 (0.26, 0.19)
5 (0.08, 0.41)
6 (0.45, 0.30)

Πίνακας απόστασης

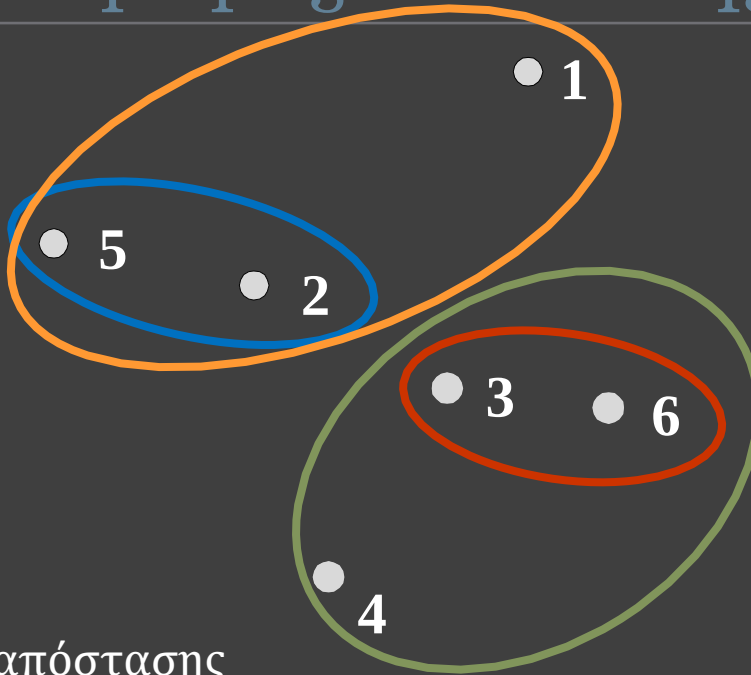
	p1	p2 ∪ p5	p3 ∪ p6 ∪ p4
p1	0.00	0.34	0.37
p2 ∪ p5	0.34	0.00	0.39
p3 ∪ p6 ∪ p4	0.37	0.39	0.00

ΒΗΜΑ 1: συγχωνεύω τις συστάδες με τη μικρότερη απόσταση (ΠΑΝΤΑ)

ΒΗΜΑ 2: για να ενημερώσω τον πίνακα: κριτήριο υπολογισμού απόστασης μεταξύ των συστάδων → **MAX**

Υπολογισμός απόστασης μεταξύ συστάδων

MAX



1 (0.4, 0.53)

2 (0.22, 0.38)

3 (0.35, 0.32)

4 (0.26, 0.19)

5 (0.08, 0.41)

6 (0.45, 0.30)

Πίνακας απόστασης

	p1 ∪ p2 ∪ p5	p3 ∪ p6 ∪ p4
p1 ∪ p2 ∪ p5	0.00	0.39
p3 ∪ p6 ∪ p4	0.39	0.00

Υπολογισμός απόστασης μεταξύ συστάδων

MAX



- 1 (0.4, 0.53)
- 2 (0.22, 0.38)
- 3 (0.35, 0.32)
- 4 (0.26, 0.19)
- 5 (0.08, 0.41)
- 6 (0.45, 0.30)

Πίνακας απόστασης

	$p1 \cup p2 \cup p5 \cup$ $p3 \cup p6 \cup p4$
$p1 \cup p2 \cup p5 \cup$ $p3 \cup p6 \cup p4$	0.00

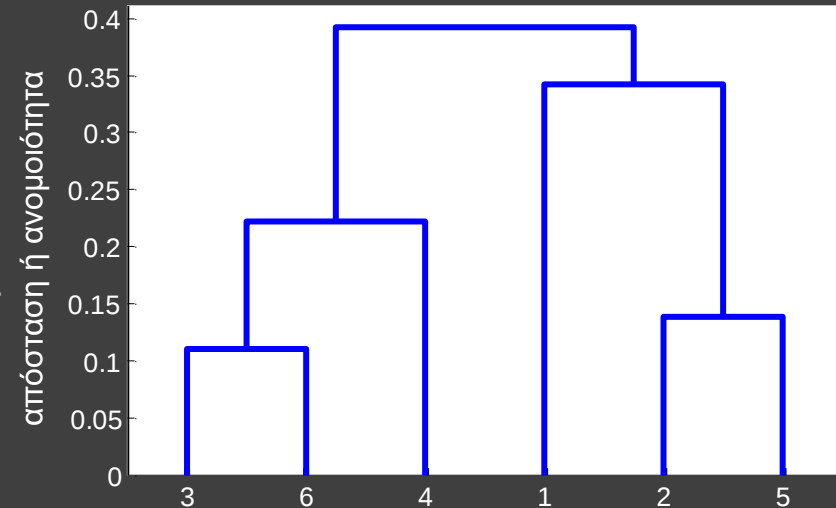
Υπολογισμός απόστασης μεταξύ συστάδων



Σειρά συγχωνεύσεων:

1. το σημείο 3 με το σημείο 6
2. το σημείο 2 με το σημείο 5
3. η ομάδα 3-6 με το σημείο 4
4. η ομάδα 2-5 με το σημείο 1
5. η ομάδες 2-5-1 με την ομάδα 3-4-6

σε αυτό το
σημείο
διαφοροποι-
είται από
τον MIN



Matlab: hierarchical clustering

4 βασικές συναρτήσεις:

<code>pdist()</code>	→	υπολογίζει την απόσταση κάθε σημείου με όλα τα υπόλοιπα
<code>linkage()</code>	→	υπολογίζει ποια σημεία/ομάδες συγχωνεύονται σε clusters
<code>dendrogram()</code>	→	δημιουργεί γράφημα με το δενδρόγραμμα
<code>cluster()</code>	→	“κόβει” το δενδρόγραμμα στο ύψος που ορίζουμε και δημιουργεί τα ανάλογα clusters

Matlab: hierarchical clustering

```
D = pdist(X, 'euclidean');
```

Υπολογίζει την απόσταση κάθε σημείου του πίνακα **X** με όλα τα υπόλοιπα, χρησιμοποιώντας την Ευκλείδεια απόσταση. Στον πίνακα **D** επιστρέφει μια τιμή απόστασης για κάθε ζευγάρι σημείων.

Μπορεί να χρησιμοποιήσει και άλλες αποστάσεις:
'squaredeuclidean', 'mahalanobis', 'minkowski', κ.λπ.

αναλυτικά: <https://www.mathworks.com/help/stats/pdist.html>

Matlab: hierarchical clustering

```
D = pdist(X, 'euclidean');
```

Υπολογίζει την απόσταση κάθε σημείου του πίνακα **X** με όλα τα υπόλοιπα, χρησιμοποιώντας την Ευκλείδεια απόσταση. Στον πίνακα **D** επιστρέφει μια τιμή απόστασης για κάθε ζευγάρι σημείων.

Μπορεί να χρησιμοποιήσει και άλλες αποστάσεις:
'squaredeuclidean', 'mahalanobis', 'minkowski', κ.λπ.

Για να μετατρέψω το διάνυσμα σε πίνακα (τετραγωνικό και συμμετρικό), χρησιμοποιώ τη `squareform()`. Αναλυτικά:
<https://www.mathworks.com/help/stats/squareform.html>

Matlab: hierarchical clustering - Παράδειγμα

```
% load data
load kmeansdata

% calculate distances between points
Distances = pdist(X, 'euclidean');
DistancesMatrix = squareform(Distances);
```

Matlab: hierarchical clustering - Παράδειγμα

```
% load data
load kmeansdata

% calculate distances between points
Distances = pdist(X, 'euclidean');
DistancesMatrix = squareform(Distances);
```

```
>> Distances  διαστάσεις: 1×156520
ans =
Columns 1 through 5
    1.8601    2.2633    2.1311    1.9134    2.4935
Columns 6 through 10
    1.3210    1.6316    0.7948    0.9172    2.3388
Columns 11 through 15
    1.2830    4.3727    0.8105    2.2500    2.9281
Columns 16 through 20
    0.3523    1.1546    2.7091    1.0528    0.8950
Columns 21 through 25
    2.9717    1.4166    2.2765    5.2127    4.3449
Columns 26 through 30
    2.3476    0.3692    2.6281    1.5837    4.6907
Columns 31 through 35
    2.2726    2.4880    4.2682    2.4821    2.5622
```

Matlab: hierarchical clustering - Παράδειγμα

```
% load data
load kmeansdata

% calculate distances between points
Distances = pdist(X, 'euclidean');
DistancesMatrix = squareform(Distances);
```

```
>> Distances
```

διαστάσεις: 1×156520

```
ans =
```

```
Columns 1 through 5
```

```
1.8601    2.2633    2.1311    1.9134
```

```
Columns 6 through 10
```

```
1.3210    1.6316    0.7948    0.9172
```

```
Columns 11 through 15
```

```
1.2830    4.3727    0.8105    2.2500
```

```
Columns 16 through 20
```

```
0.3523    1.1546    2.7091    1.0528
```

```
Columns 21 through 25
```

```
2.9717    1.4166    2.2765    5.2127
```

```
Columns 26 through 30
```

```
2.3476    0.3692    2.6281    1.5837
```

```
Columns 31 through 35
```

```
0.2720    0.4000    0.2600    0.4000
```

```
>> DistancesMatrix
```

διαστάσεις: 560×560

```
ans =
```

```
Columns 1 through 4
```

```
0    1.8601    2.2633    2.1311
```

```
1.8601    0    1.3514    1.3886
```

```
2.2633    1.3514    0    2.3089
```

```
2.1311    1.3886    2.3089    0
```

```
1.9134    1.5008    0.5939    2.4767
```

```
2.4935    0.8238    1.1749    1.5307
```

```
1.3210    2.4813    2.7679    2.0549
```

```
1.6316    2.2186    2.7723    1.6302
```

```
0.7948    1.2375    1.8634    1.8586
```

```
0.9172    2.3289    2.2238    2.8117
```

```
4.6907
```

Matlab: hierarchical clustering

```
ClusterTree = linkage(X);
```

Υπολογίζει ποια σημεία/ομάδες του πίνακα **X** συγχωνεύονται σε clusters. Ο πίνακας **ClusterTree** που επιστρέφει, έχει 3 στήλες:

Οι δύο πρώτες στήλες είναι το ζευγάρι των σημείων ή clusters που συγχωνεύονται και η τρίτη στήλη είναι η μεταξύ τους απόσταση.

- Στις 2 πρώτες στήλες εμφανίζονται τιμές μεγαλύτερες από το σύνολο των σημείων που έχουμε. Αυτό συμβαίνει επειδή εξορισμού ονομάζει κάθε νέα συγχώνευση με έναν καινούργιο αριθμό. Δηλαδή:
σημείο 3 & σημείο 5 → cluster 11
cluster 11 και σημείο 1 → cluster 12 κτλ.

αναλυτικά: <https://www.mathworks.com/help/stats/linkage.html>

Matlab: hierarchical clustering - Παράδειγμα

```
% create cluster tree  
ClusterTree = linkage(X);
```

```
>> ClusterTree  
ans =  
  
    374    417    0.13007  
    300    385    0.13271  
    366    383    0.15739  
    340    349    0.17503  
    101    177    0.18698  
    113    170    0.19221  
    471    549    0.20553  
    157    278    0.20821  
     91    201    0.21755  
    297    318    0.21903  
    436    541    0.22355  
    145    228    0.22581  
    103    249    0.24169  
     88    190    0.24179  
    319    322    0.24365  
    284    382    0.25085  
     90    174    0.2519  
     93    176    0.26212
```

Οι δύο πρώτες στήλες είναι το ζευγάρι των σημείων ή clusters που συγχωνεύονται

η τρίτη στήλη είναι η μεταξύ τους απόσταση

Matlab: hierarchical clustering

```
ClusterTree = linkage(X,method);
```

Χρησιμοποιεί τη μέθοδο που προσδιορίζεται στο `method` για κριτήριο επιλογής της απόστασης:

<code>'single'</code>	:	μικρότερη απόσταση (MIN)
<code>'complete'</code>	:	μεγαλύτερη απόσταση (MAX)
<code>'centroid'</code>	:	απόσταση μεταξύ των κέντρων
<code>'ward'</code>	:	μέθοδος του Ward

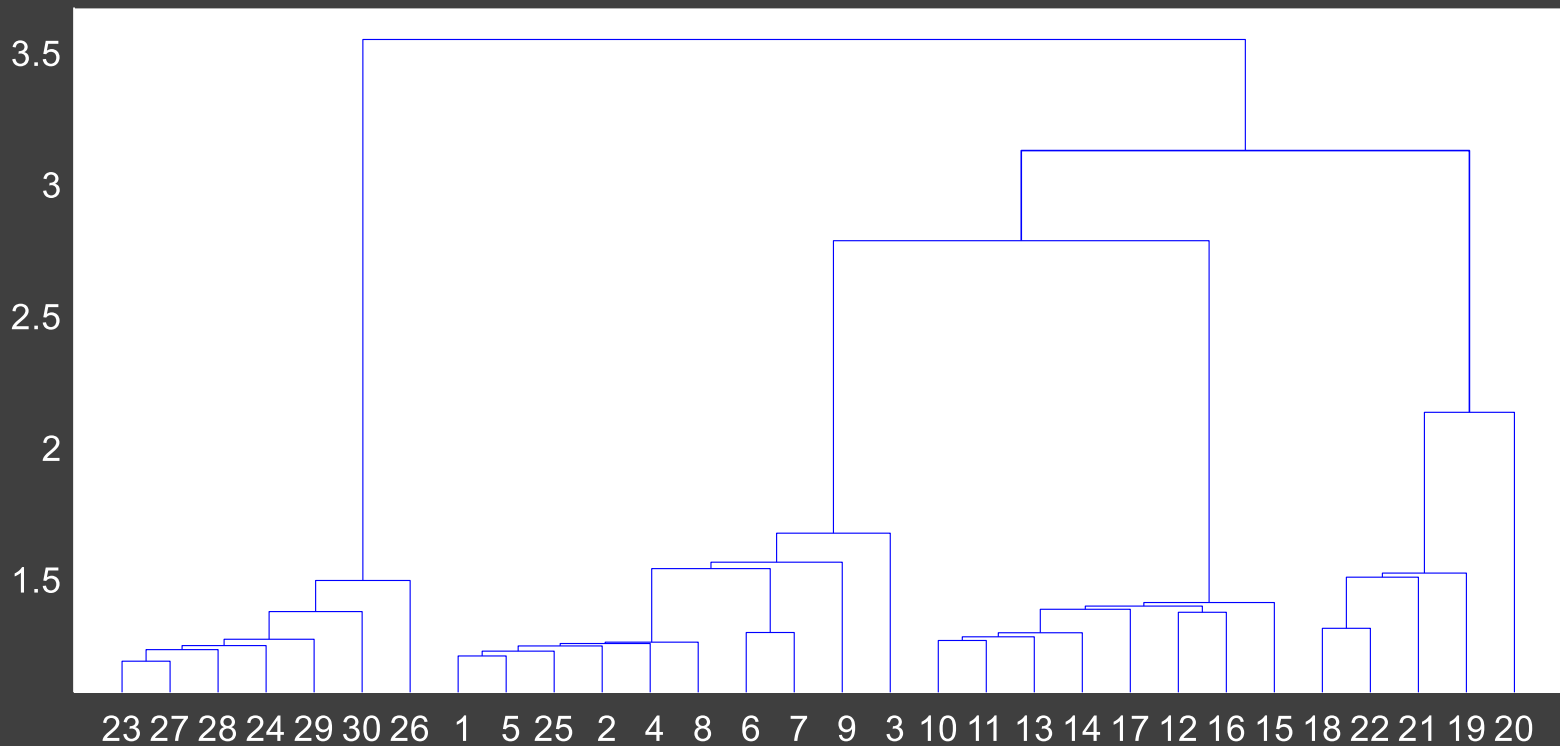
π.χ.

```
ClusterTree = linkage(X, 'single');
```

Matlab: hierarchical clustering

```
dendrogram(ClusterTree);
```

Δημιουργεί γράφημα με το δενδρόγραμμα



αναλυτικά:

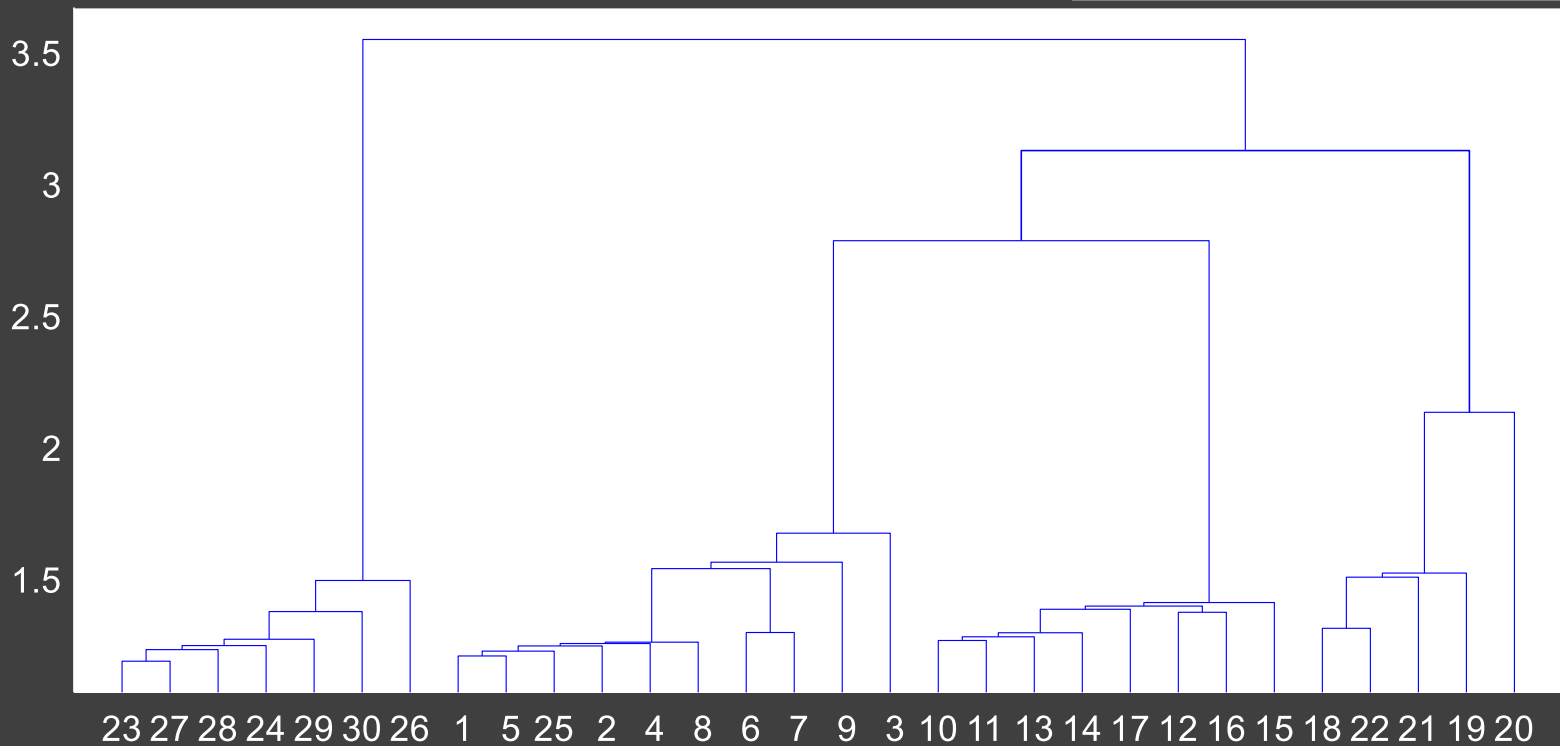
<https://www.mathworks.com/help/stats/dendrogram.html>

Matlab: hierarchical clustering

```
dendrogram(ClusterTree);
```

Δημιουργεί γράφημα με το δενδρόγραμμα

Αν υπάρχουν περισσότερα από 30 σημεία δεδομένων, δεν εμφανίζονται τα χαμηλότερα κλαδιά, το δενδρόγραμμα “κόβεται” στα 30 φύλλα



αναλυτικά:

<https://www.mathworks.com/help/stats/dendrogram.html>

Matlab: hierarchical clustering

```
Clusters = cluster(ClusterTree, 'Cutoff', a);
```

“Κόβει” το δενδρόγραμμα στο σημείο που ορίζουμε (a) και δημιουργεί τα ανάλογα clusters.

Εκτός από το a παίρνει ως είσοδο και το αποτέλεσμα της συνάρτησης `linkage()`.

Επιστρέφει ένα διάνυσμα με τις συστάδες στις οποίες ανήκουν τα σημεία των αρχικών δεδομένων, αντίστοιχο με το διάνυσμα που επιστρέφει η συνάρτηση `kmeans()`.

αναλυτικά: <https://www.mathworks.com/help/stats/cluster.html>

Matlab: hierarchical clustering

```
Clusters =  
cluster(ClusterTree, 'Cutoff', a, 'Criterion',  
criterion);
```

Το σημείο *a* μπορεί να είναι είτε το ύψος (στο δενδρόγραμμα) στο οποίο θέλουμε να “κόψει”, είτε ο συντελεστής αστάθειας (inconsistency coefficient), ο οποίος συγκρίνει το ύψος κάθε σύνδεσης του δένδρου με το μέσο ύψος των γειτονικών συνδέσεων. Γενικά μικρότερος συντελεστής σημαίνει καλύτερο clustering.

Ποιο από τα δυο κριτήρια να χρησιμοποιήσει το ορίζουμε στο `criterion`: `'inconsistent'` ή `'distance'`

π.χ.:

```
Clusters =  
cluster(ClusterTree, 'Cutoff', 2.5, 'Criterion', 'distance');
```

Matlab: hierarchical clustering - Παράδειγμα

```
% create clusters  
Clusters = cluster(ClusterTree, 'Cutoff', 2.5, 'Criterion', 'distance');
```

```
>> Clusters  
ans =  
    4  
    4  
    4  
    4  
    4  
    1  
    1  
    1  
    1  
    2  
    2  
    2  
    2  
    2  
    2  
    2  
    2  
    3  
    3
```