

# Introduction au Machine Learning

## Classification non supervisée

**Emmanuel Dellandréa**

**[emmanuel.dellandrea@ec-lyon.fr](mailto:emmanuel.dellandrea@ec-lyon.fr)**

Version du 06/01/2025



# Classification non supervisée

- Introduction
- Terminologie
- Les centre mobiles
- Classification ascendante hiérarchique (CAH)
- Classification mixte : CAH + centres mobiles

# Introduction

- Les techniques de classification non supervisée sont destinées à produire des groupements de lignes ou de colonnes d'un tableau
- Le contexte général d'utilisation est l'analyse d'un grand tableau de données

# Introduction

- On suppose que certains regroupements existent dans les données
- On cherche donc à les mettre en évidence par des représentations des classes d'éléments sous forme de partitions (ou hiérarchies de partitions)

# Introduction

- Il existe plusieurs familles d'algorithmes de classification
  - Les algorithmes construisant directement des partitions comme les méthodes d'agrégation autour des centres mobiles
  - Les algorithmes ascendants qui procèdent à la construction de classes par agglomérations successives des objets deux à deux, et qui fournissent une hiérarchie de partitions des objets
  - Les algorithmes descendants qui procèdent par dichotomies successives de l'ensemble des objets et qui peuvent également fournir une hiérarchie de partitions

# Introduction

- Ces techniques présentent des avantages différents et peuvent être utilisées conjointement sous forme de méthodes hybrides
- Un des avantages des méthodes de classification est de mettre en évidence des classes regroupant des éléments partageant des caractéristiques communes

# Classifications non supervisées

- Introduction
- ➔ ▪ Terminologie
- Les centre mobiles
- Classification ascendante hiérarchique (CAH)
- Classification mixte : CAH + centres mobiles

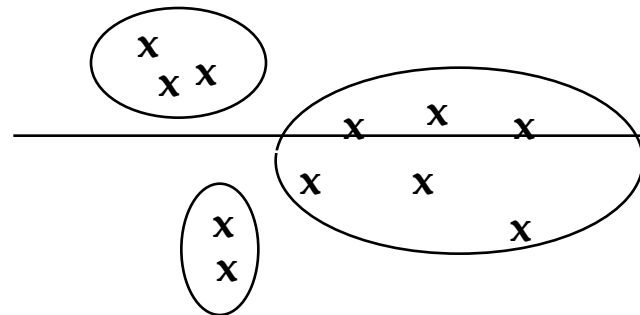
# Partitions

- Ensemble de parties non vides de l'ensemble des individus, dont les intersections deux à deux sont vides, et dont la réunion forme l'univers

$$\forall l \in \{1, 2, \dots, k\}, P_l \neq \emptyset$$

$$\forall l, m \in \{1, 2, \dots, k\}, l \neq m, P_l \cap P_m = \emptyset$$

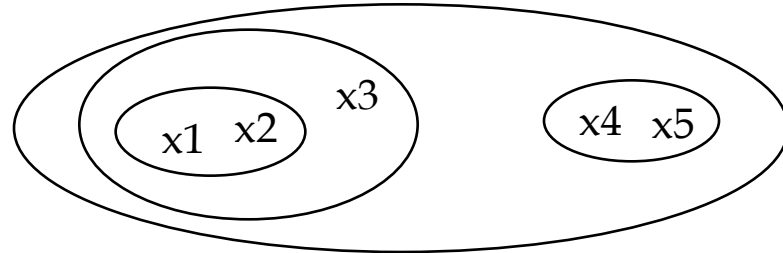
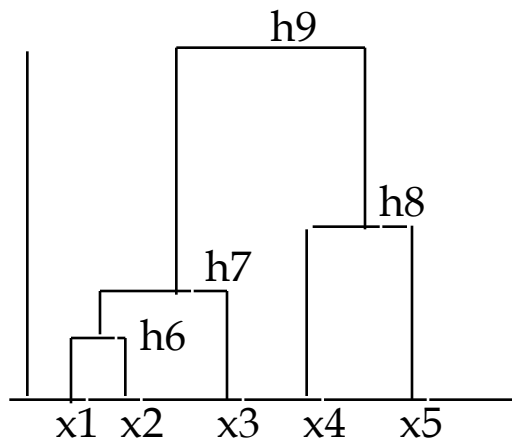
$$\bigcup_{l=1}^k P_l = \Omega$$





# Hiérarchies

- Ensemble de partitions emboîtées
- Hiérarchie = arbre



# Classifications non supervisées

- Introduction
- Terminologie
- ➔ ▪ Les centre mobiles
- Classification ascendante hiérarchique (CAH)
- Classification mixte : CAH + centres mobiles

# Les centres mobiles

- C'est la méthode de partitionnement la mieux adaptée aux vastes ensembles de données
- Elle est utilisée comme
  - Technique de description et d'analyse
  - Technique de réduction (mise en évidence d'éléments « prototypes » représentatif des classes)

# Les centres mobiles

## ■ Bases théoriques de l'algorithme

- Soit un ensemble  $I$  de  $n$  individus à partitionner, caractérisés par  $p$  variables
- On suppose que l'espace  $R^p$  est muni d'une distance appropriée, nommée  $d$  (souvent distance euclidienne usuelle)
- On désire constituer un maximum de  $q$  classes
- L'algorithme procède en  $m$  étapes

# Les centres mobiles

- Bases théoriques de l'algorithme

- Etape 0 :

- On détermine  $q$  centres provisoires de classes (par exemple par un tirage aléatoire des individus)

$$\{C_1^0, \dots, C_k^0, \dots, C_q^0\}$$

- Ces  $q$  centres induisent une première partition  $P^0$  de l'ensemble des  $I$  individus en  $q$  classes

$$\{I_1^0, \dots, I_k^0, \dots, I_q^0\}$$

- L'individu  $i$  appartient à la classe  $I_k^0$  s'il est plus proche de  $C_k^0$  que de tous les autres centres

# Les centres mobiles

- Bases théoriques de l'algorithme
  - Etape 1 :
    - On détermine  $q$  nouveaux centres de classes  $\{C_I^1, \dots, C_k^1, \dots, C_q^1\}$  en prenant les centres de gravité des classes qui viennent d'être obtenues :  $\{I_I^0, \dots, I_k^0, \dots, I_q^0\}$
    - Ces nouveaux centres induisent une nouvelle partition  $P^1$  de  $I$  construite selon la même règle que pour  $P^0$  :  
 $\{I_I^1, \dots, I_k^1, \dots, I_q^1\}$

# Les centres mobiles

- Bases théoriques de l'algorithme

- Etape  $m$  :

- On détermine  $q$  nouveaux centres de classes

$\{C_I^m, \dots, C_k^m, \dots, C_q^m\}$  en prenant les centres de gravité des classes qui obtenues lors de l'étape précédente :  $\{I_I^{m-1}, \dots, I_k^{m-1}, \dots, I_q^{m-1}\}$

- Ces nouveaux centres induisent une nouvelle partition  $P^m$  de  $I$  formée des classes :

$\{I_I^m, \dots, I_k^m, \dots, I_q^m\}$

# Les centres mobiles

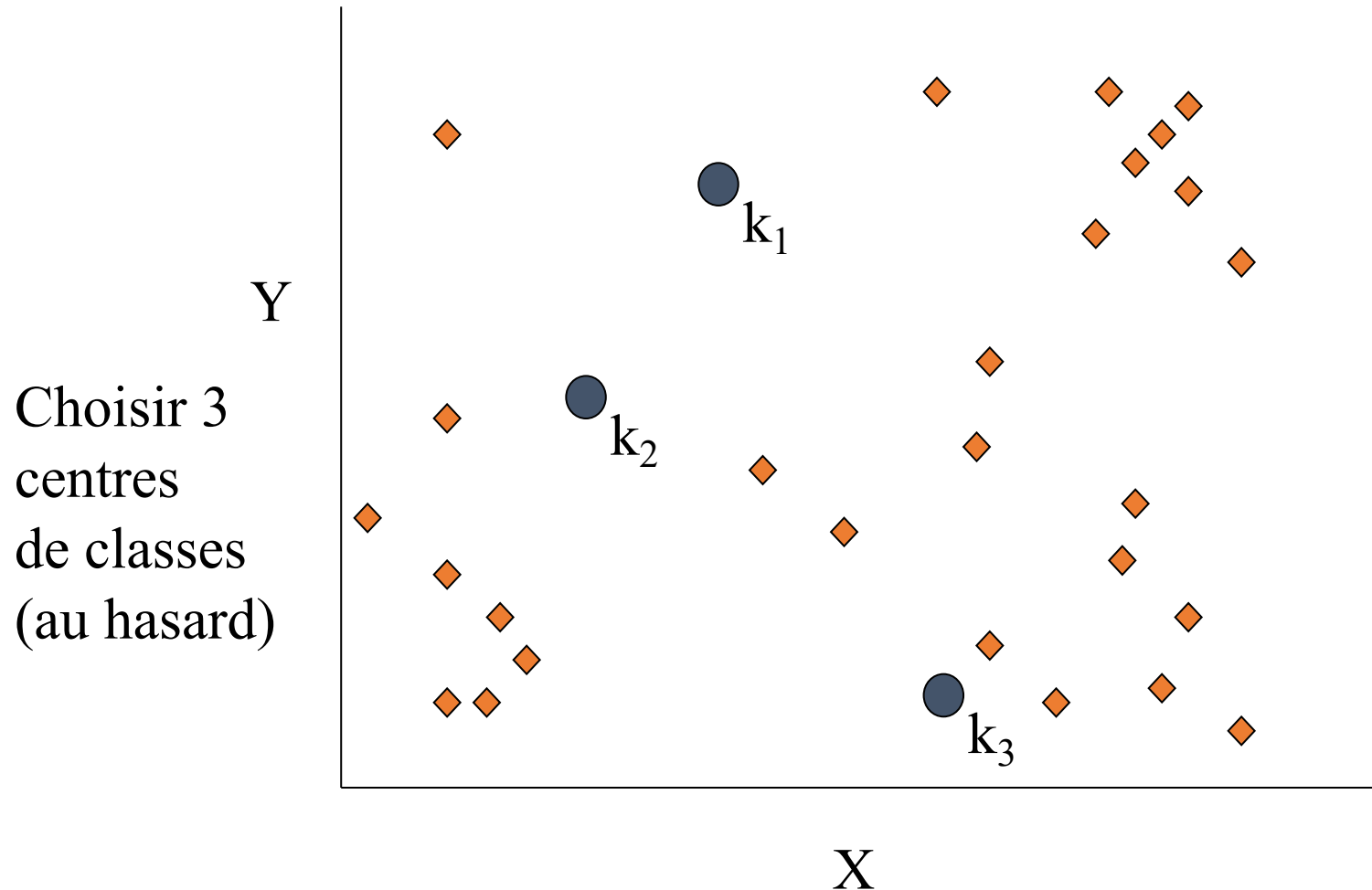
- Bases théoriques de l'algorithme
  - Le processus se stabilise nécessairement et l'algorithme s'arrête lorsque deux itérations successives conduisent à une même partition
  - Deux autres critères d'arrêt peuvent être utilisés
    - La mesure de la variance intra-classe cesse de décroître
    - Un nombre maximal d'itérations a été atteint



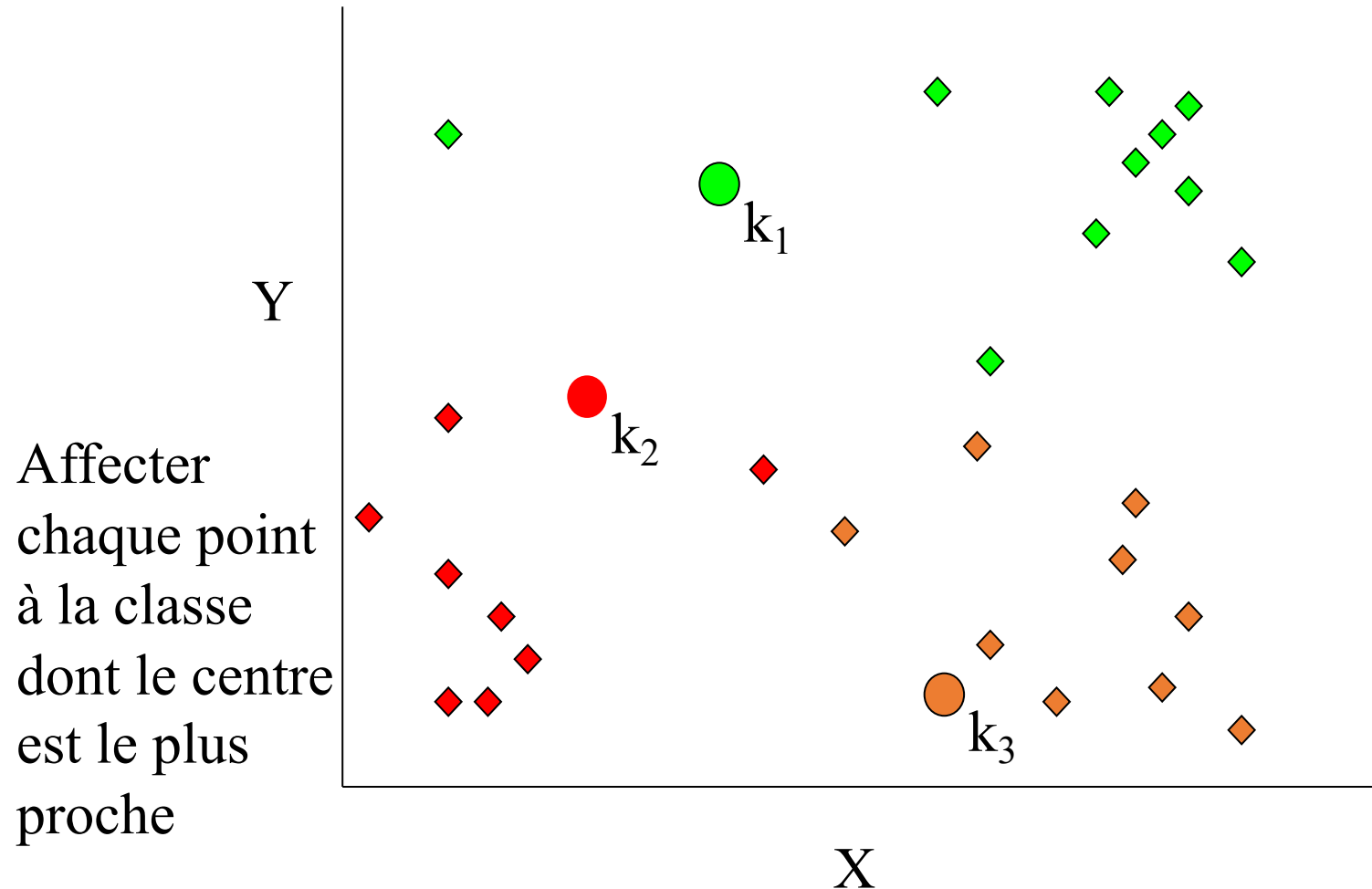
# Les centres mobiles

- Bases théoriques de l'algorithme
  - La partition obtenue dépend généralement du choix initial des centres

# Les centres mobiles : exemple

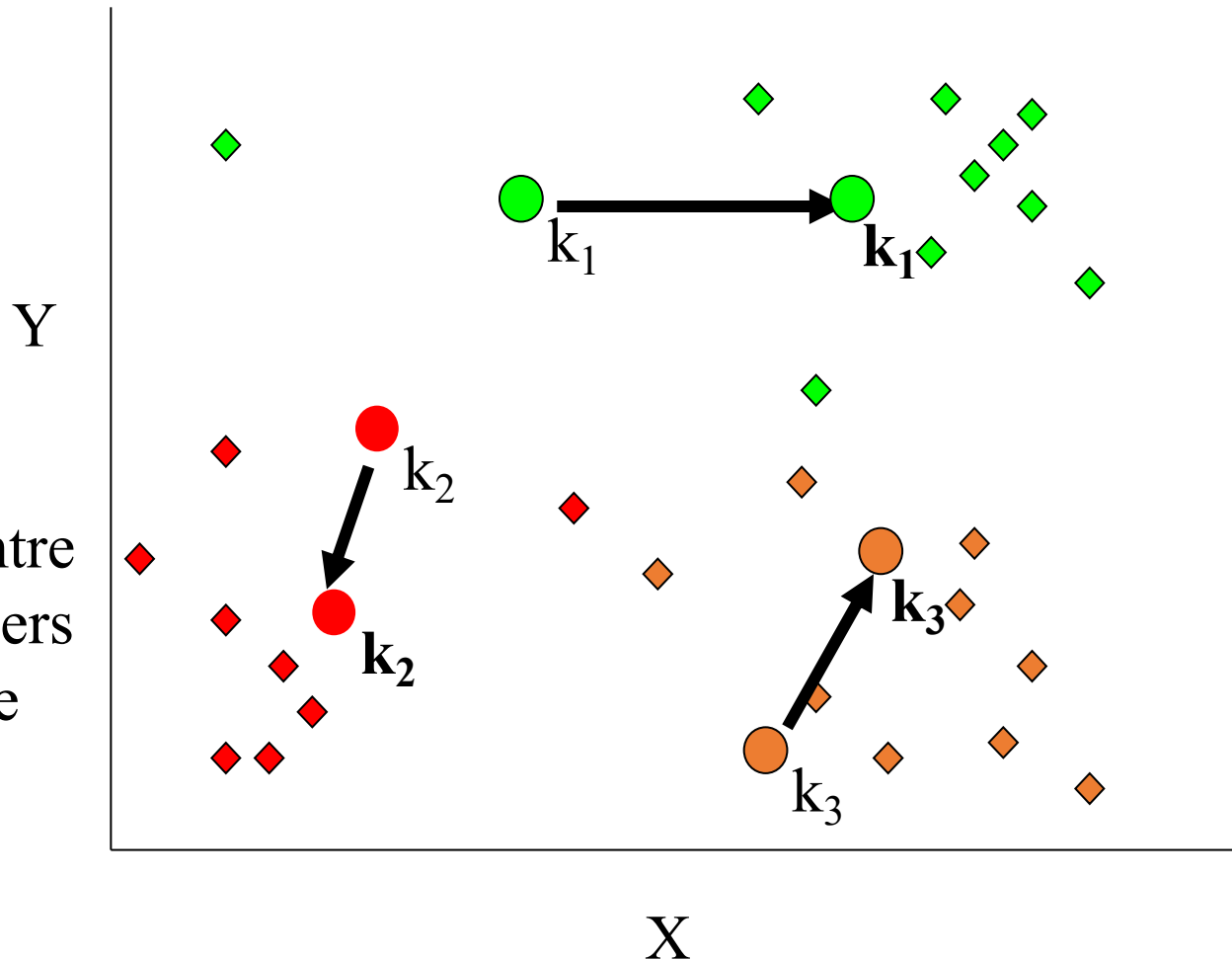


# Les centres mobiles : exemple



# Les centres mobiles : exemple

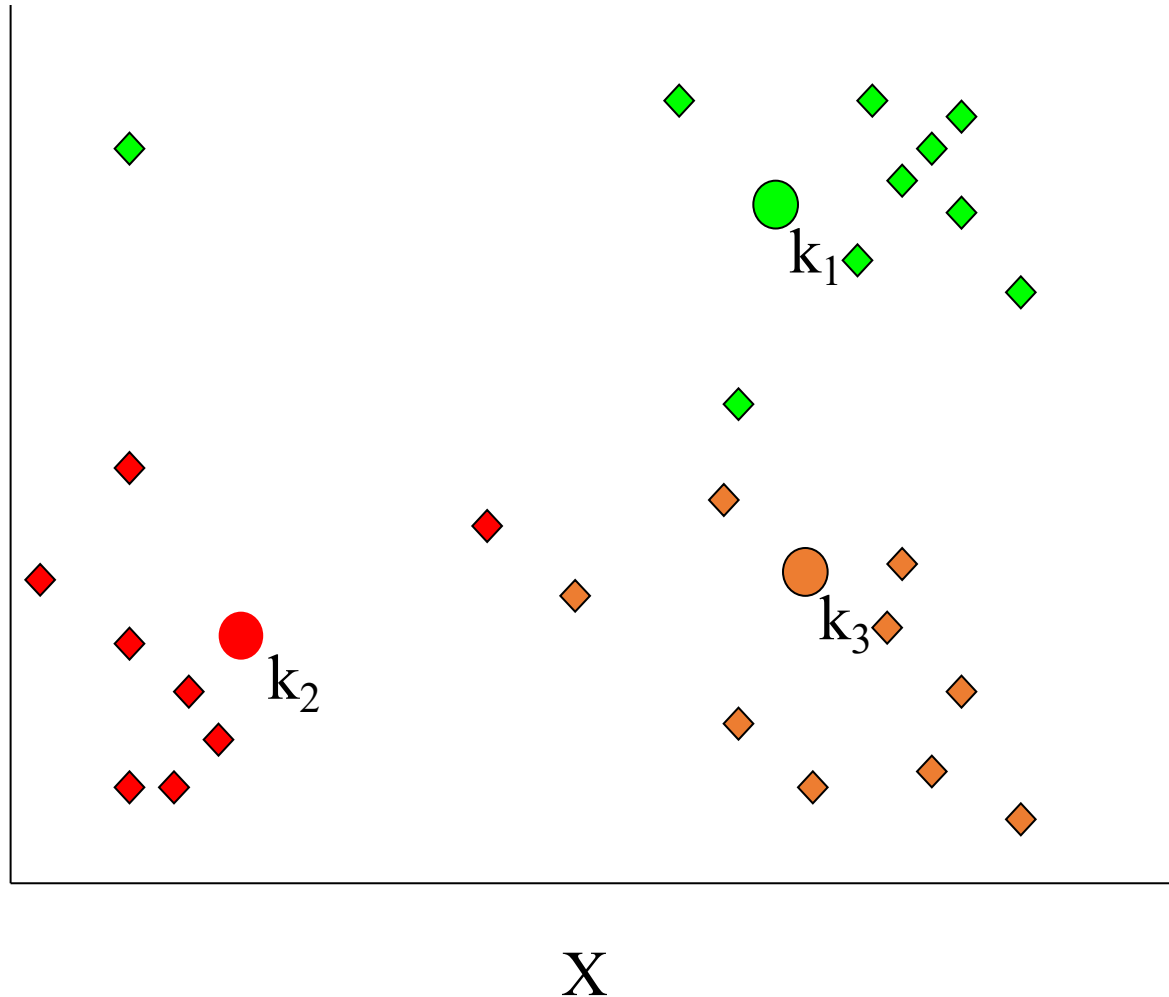
Déplacer  
chaque centre  
de classe vers  
la moyenne  
de chaque  
classe



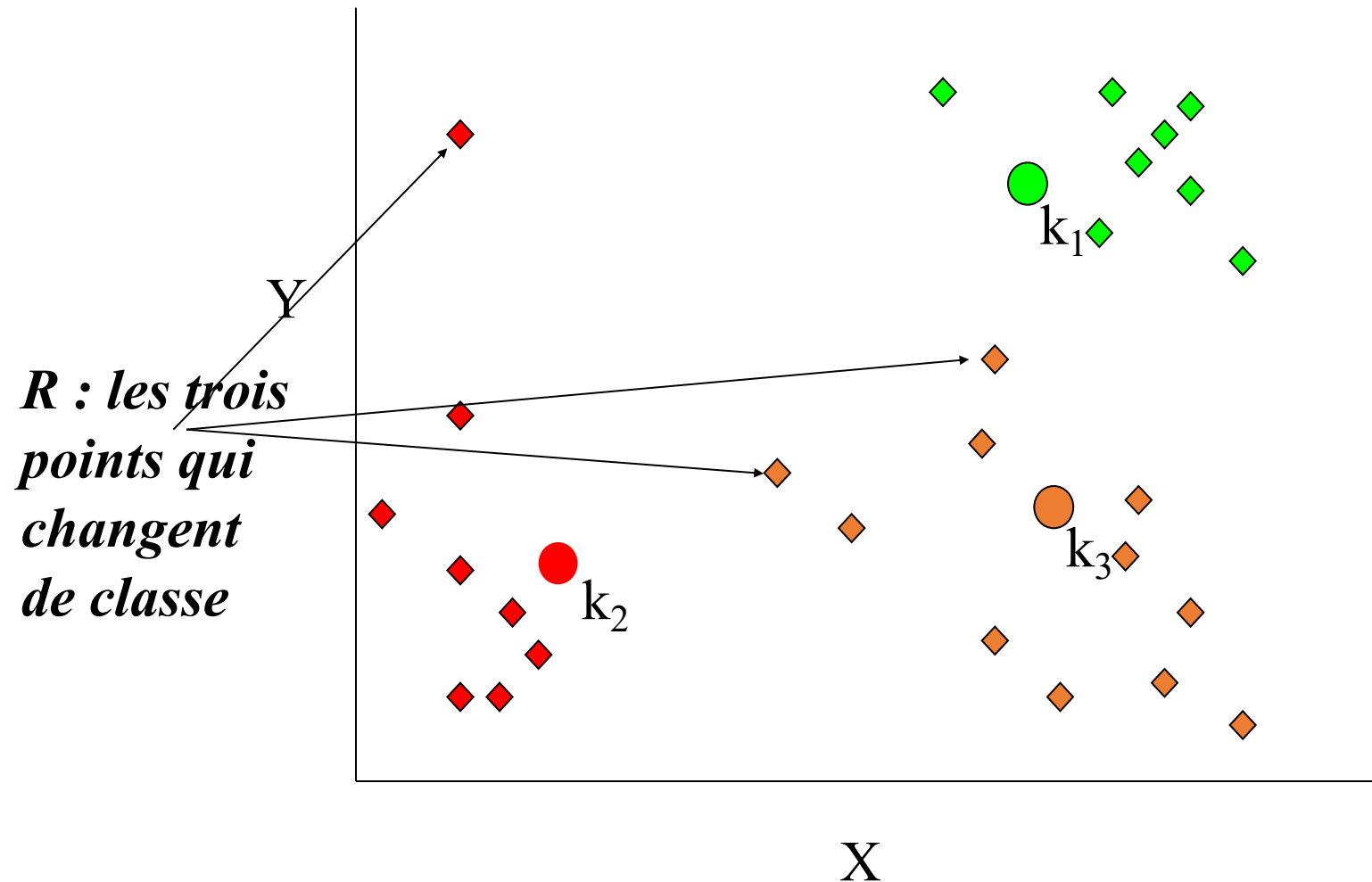
# Les centres mobiles : exemple

Réaffecter les points qui sont plus proches du centre d'une autre classe  $Y$

*Q : Quels sont les points qui changent de classe?*



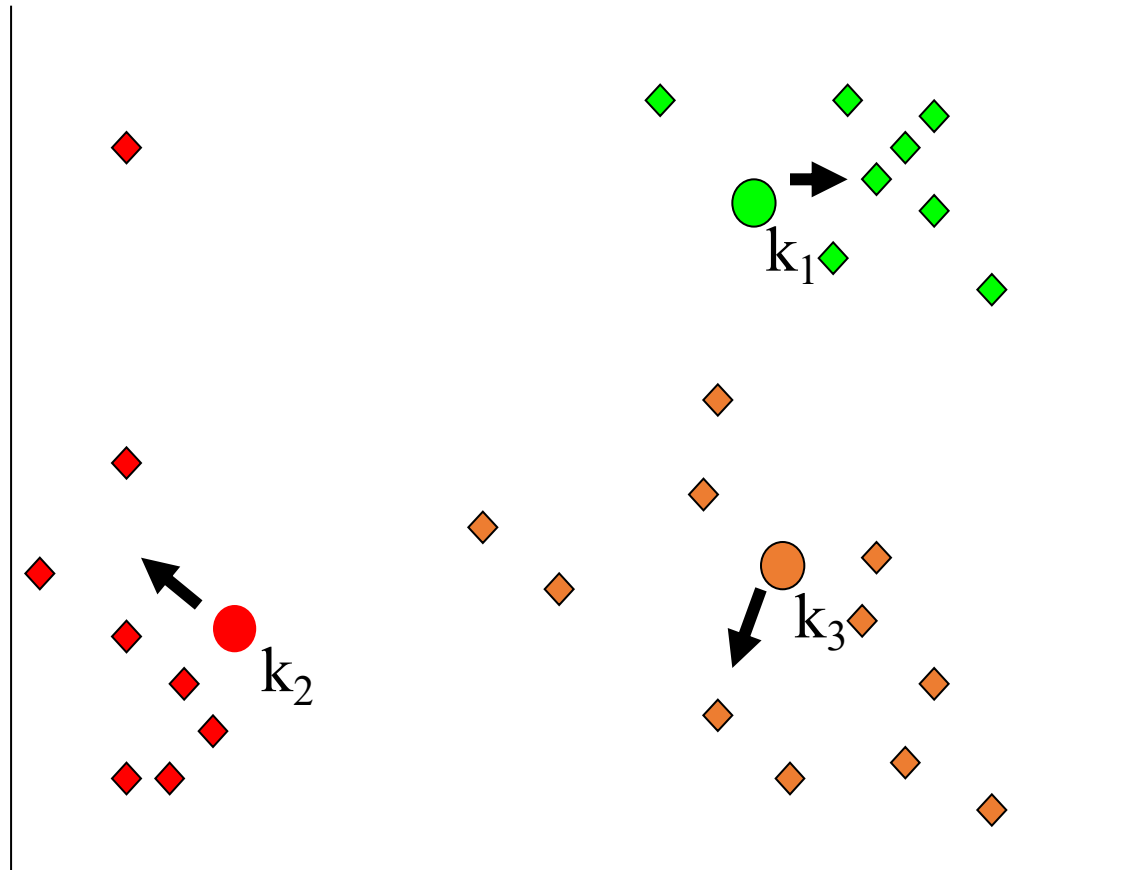
# Les centres mobiles : exemple



# Les centres mobiles : exemple

Re-calculer  
les moyennes  
des classes

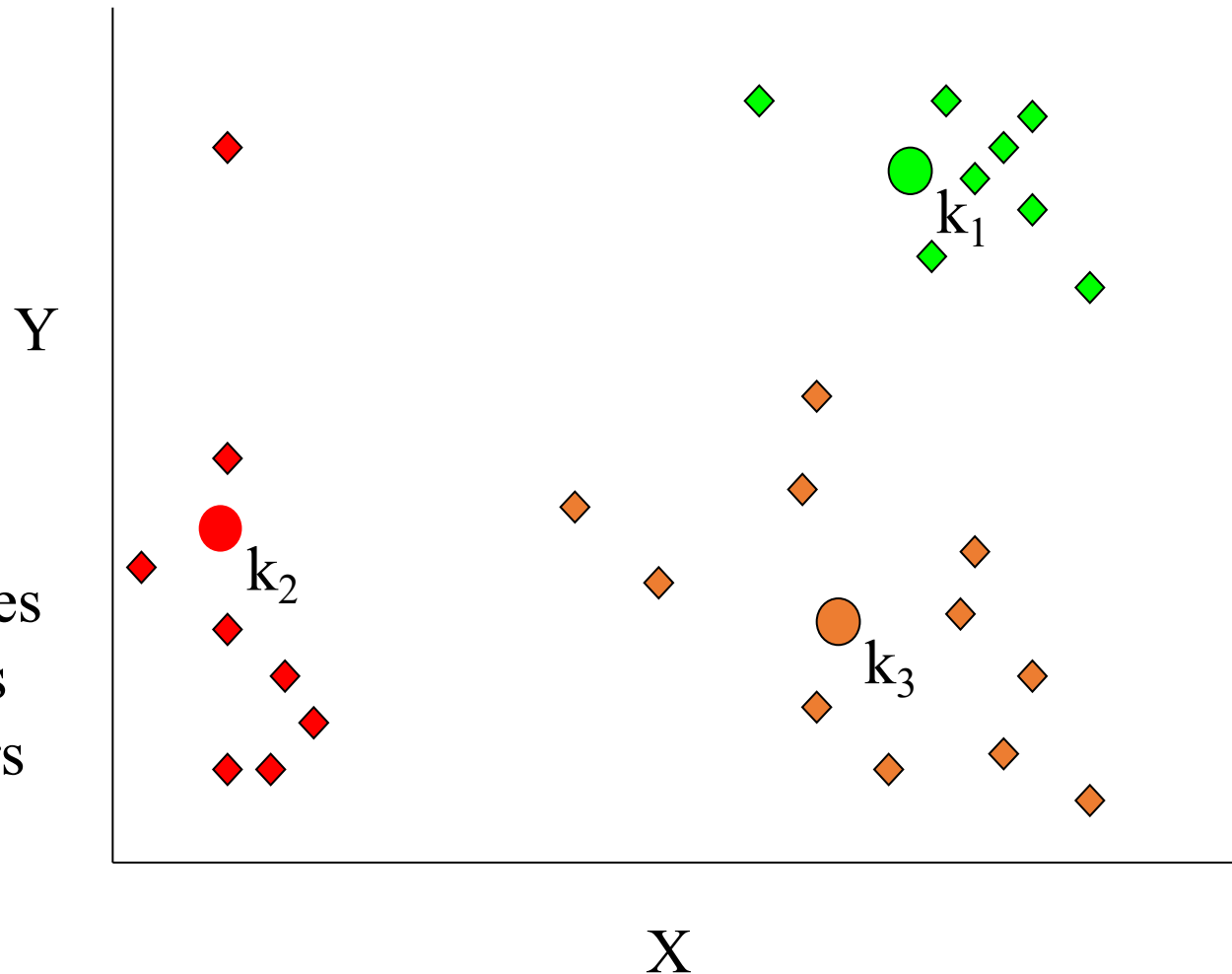
Y



X

# Les centres mobiles : exemple

Déplacer les  
centres des  
classes vers  
les  
moyennes





# Les centres mobiles

- Il existe plusieurs variantes de l'algorithme des centres mobiles, dépendant du choix
  - Des distances utilisées
  - Des règles d'affectation
- Exemple : les nuées dynamiques
  - Les classes ne sont pas caractérisées par un centre de gravité, mais par un certain nombre d'individus qui constituent un « noyau »
  - Ce « noyau » a pour certaines utilisations un meilleur pouvoir descriptif et discriminant qu'un centre ponctuel

# Classifications non supervisées

- Introduction
- Terminologie
- Les centre mobiles
- ➔ ▪ Classification ascendante hiérarchique (CAH)
- Classification mixte : CAH + centres mobiles

# Classification ascendante hiérarchique

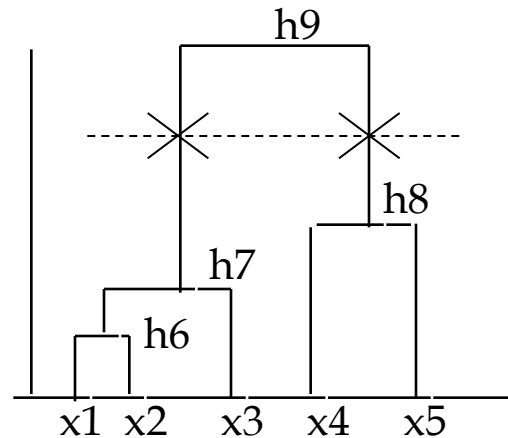
- Le principe de cet algorithme est de créer à chaque étape une partition obtenue en agrégeant deux à deux les éléments les plus proches
- Un élément peut désigner un individu, ou un regroupement d'individus généré par l'algorithme
- Il existe différentes manières de considérer le nouveau couple d'éléments agrégés, d'où un nombre important de variantes de cette technique

# Classification ascendante hiérarchique

- L'algorithme ne fournit pas une partition en  $q$  classes d'un ensemble de  $n$  objets, mais une hiérarchie de partitions, se présentant sous la forme d'arbres (appelés dendrogrammes) et contenant  $n-1$  partitions
- L'intérêt de ces arbres est qu'ils peuvent donner une idée du nombre de classes existant effectivement dans la population
- Chaque coupure de l'arbre fournit une partition : plus on coupe haut, plus le nombre de classes est faible, et moins les classes sont homogènes

# Classification ascendante hiérarchique

- Dendrogramme



# Classification ascendante hiérarchique

## ■ Critères d'agrégation

- On suppose que l'ensemble des individus est muni d'une distance  $d$
- Lorsqu'un groupe d'individus est constitué, il convient de se demander sur quelle base on peut calculer une distance entre un individu et un groupe, ainsi qu'une distance entre deux groupes
- Ceci revient à définir une stratégie de regroupement, appelée critères d'agrégation
- La distance entre groupes pourra généralement se calculer à partir des distances des différents éléments impliqués dans le groupe

# Classification ascendante hiérarchique

- Critères d'agrégation
  - Les critères les plus usuels sont les suivants
    - Critère du saut minimal
    - Critère du saut maximal
    - Critère de la distance moyenne
    - Critère de la distance moyenne pondérée
    - Critère de Ward

# Classification ascendante hiérarchique

- Critères d'agrégation
  - Critère du saut minimal
    - Soit  $x, y$  et  $z$  trois objets
    - Les objets  $x$  et  $y$  sont regroupés en un seul élément noté  $h$
    - Le critère du saut minimal consiste à définir la distance de ce groupe à  $z$  par la plus petite distance des éléments de  $h$  à  $z$   
$$d(h,z) = \text{Min}\{ d(x,z), d(y,z) \}$$



# Classification ascendante hiérarchique

- Critères d'agrégation

- Critère du saut maximal

- Soit  $x, y$  et  $z$  trois objets
      - Les objets  $x$  et  $y$  sont regroupés en un seul élément noté  $h$
      - Le critère du saut maximal consiste à définir la distance de ce groupe à  $z$  par la plus grande distance des éléments de  $h$  à  $z$
- $$d(h,z) = \text{Max}\{ d(x,z), d(y,z) \}$$

# Classification ascendante hiérarchique

- Critères d'agrégation

- Critère de la distance moyenne

- Soit  $x, y$  et  $z$  trois objets
    - Les objets  $x$  et  $y$  sont regroupés en un seul élément noté  $h$
    - Le critère de la distance moyenne consiste à définir la distance de ce groupe à  $z$  par la moyenne des distances des éléments de  $h$  à  $z$

$$d(h,z) = ( d(x,z) + d(y,z) ) / 2$$

# Classification ascendante hiérarchique

- Critères d'agrégation
  - Critère de la distance moyenne pondérée
    - Soit  $x$ ,  $y$  et  $z$  trois objets,  $x$  contenant  $n_x$  éléments et  $y$  en contenant  $n_y$
    - Les objets  $x$  et  $y$  sont regroupés en un seul élément noté  $h$  contenant donc  $(n_x + n_y)$  éléments
    - Le critère de la distance moyenne pondérée consiste à définir la distance de ce groupe à  $z$  par la distance suivante
$$d(h,z) = (n_x d(x,z) + n_y d(y,z)) / (n_x + n_y)$$

# Classification ascendante hiérarchique

- Critères d'agrégation

- Critère de Ward

- Cette technique d'agrégation, basée sur la variance, cherche à optimiser à chaque étape, au sens de l'inertie, la partition obtenue par agrégation de deux éléments
    - En effet, la qualité d'une partition est intuitivement bonne si
      - À l'intérieur de chaque classe, la variance des individus qui la composent est faible pour chaque variable (inertie intra-classe)
      - La variance entre classes est importante pour chaque variable (inertie inter-classes)

# Classification ascendante hiérarchique

- Critères d'agrégation

- Critère de Ward

- Le théorème de Huygens exprime la décomposition de l'inertie totale d'une partition par :

$$\textit{Inertie totale} = \textit{Inertie inter-classes} + \textit{Inertie intra-classes}$$

- Puisque l'inertie totale est fixée par les données, cette relation montre qu'il est équivalent de rechercher une partition présentant une inertie inter-classes élevée ou une inertie intra-classe faible

# Classification ascendante hiérarchique

- Critères d'agrégation
  - Critère de Ward
    - A une étape donnée, en agrégeant deux classes, l'inertie intra-classe est nécessairement plus grande (éventuellement égale) que celle de la partition obtenue à l'étape précédente
    - L'objectif de ce critère est donc de choisir à chaque étape le regroupement de classes tel que l'augmentation de l'inertie intra-classe de la partition soit minimale

# Classification ascendante hiérarchique

- Critères d'agrégation

- Critère de Ward

- Soit  $C_i$  et  $C_j$  les centres de gravités des classes respectives  $I_i$  et  $I_j$
    - Soit  $p_i$  et  $p_j$  la somme des poids des éléments des classes  $I_i$  et  $I_j$
    - On montre que l'augmentation de l'inertie intra-classe due au regroupement des classes  $I_i$  et  $I_j$  s'écrit :

$$\delta(I_i, I_j) = \frac{p_i p_j}{p_i + p_j} d^2(C_i, C_j)$$

# Classification ascendante hiérarchique

- Critères d'agrégation

- Critère de Ward

- Cette mesure de l'augmentation de l'inertie intra-classe est le critère à minimiser à chaque étape
    - Elle montre qu'à chaque étape, on regroupe les classes
      - Proches (minimisant la distance  $d^2(C_i, C_j)$  )
      - De faibles poids (minimisant  $p_i p_j / (p_i + p_j)$  )



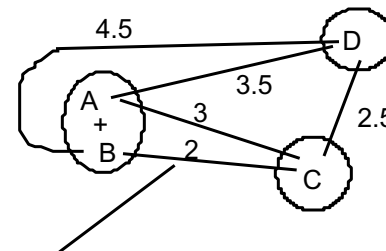
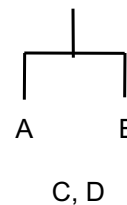
# Influence du critère d'agrégation

Tableau des distances

	A	B	C	D
A	0	1	3	3.5
B	1	0	2	4.5
C	3	2	0	2.5
D	3.5	4.5	2.5	0

$d(A,B)$  est  
la dist min

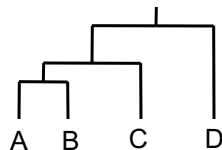
on agrège  
A avec B



Quelle distance choisir pour  $d(A+B, C)$  et pour  $d(A+B, D)$

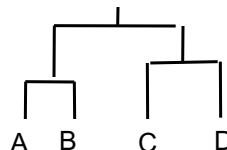
	A+B	C	D
A+B	0	2	3.5
C	2	0	2.5
D	3.5	2.5	0

Stratégie du saut minimum  
On choisit la distance Min



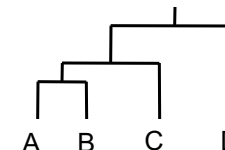
	A+B	C	D
A+B	0	3	4.5
C	3	0	2.5
D	4.5	2.5	0

Stratégie du saut maximum  
On choisit la distance Max



	A+B	C	D
A+B	0	2.5	4
C	2.5	0	2.5
D	4	2.5	0

Stratégie de la moyenne  
On choisit la distance Moyenne



# Classification ascendante hiérarchique

- **Algorithme de classification**
  - Etape 1 : il y a  $n$  éléments à classer (les  $n$  individus)
  - Etape 2 :
    - On construit la matrice des distances entre les  $n$  éléments et on cherche les deux plus proches, que l'on agrège en un nouvel élément
    - On obtient une première partition à  $n-1$  classes

# Classification ascendante hiérarchique

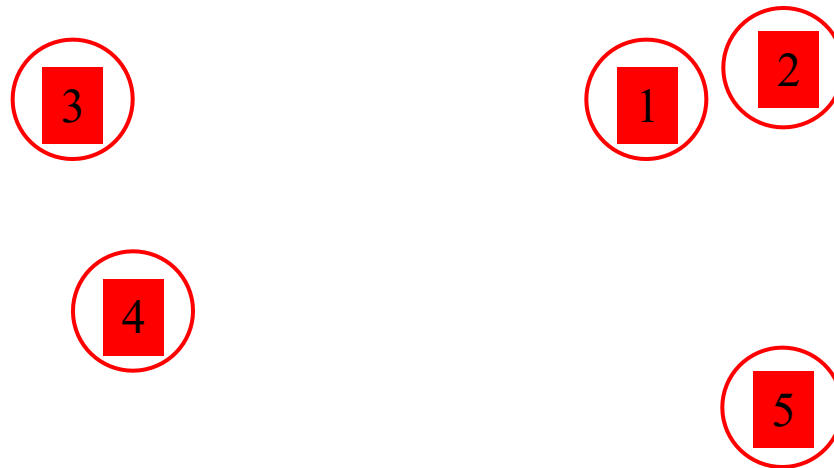
- **Algorithme de classification**
  - Etape 3 :
    - On construit une nouvelle matrice des distances, en calculant les distances, pour un critère d'agrégation donné, entre le nouvel élément et les éléments restants (les autres distances sont inchangées)
    - On cherche à nouveau les deux éléments les plus proches, que l'on agrège en un nouvel élément
    - On obtient une deuxième partition à  $n-2$  classes, qui englobe la première

# Classification ascendante hiérarchique

- **Algorithme de classification**
  - Etape  $m$  :
    - On construit une nouvelle matrice des distances
    - Le processus est réitéré jusqu'à n'avoir plus qu'un seul élément regroupant tous les objets et constituant la dernière partition
  - Les regroupements successifs peuvent être représentés par un dendrogramme

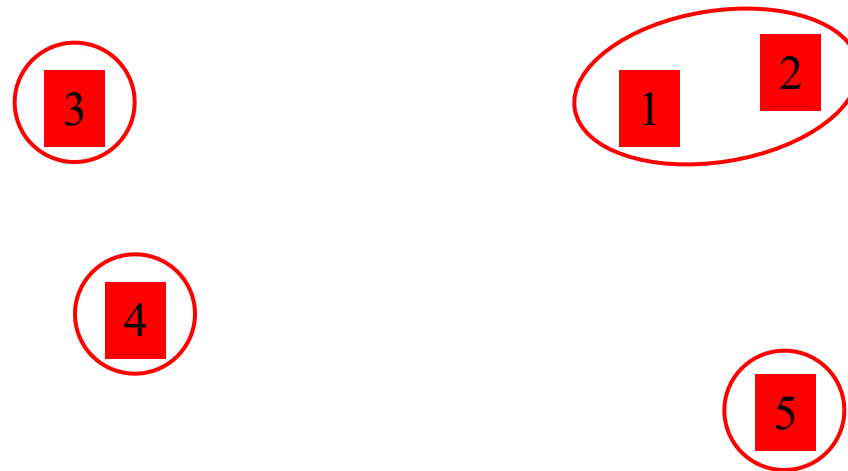
# Classification ascendante hiérarchique

- Illustration de l'agglomération progressive des éléments



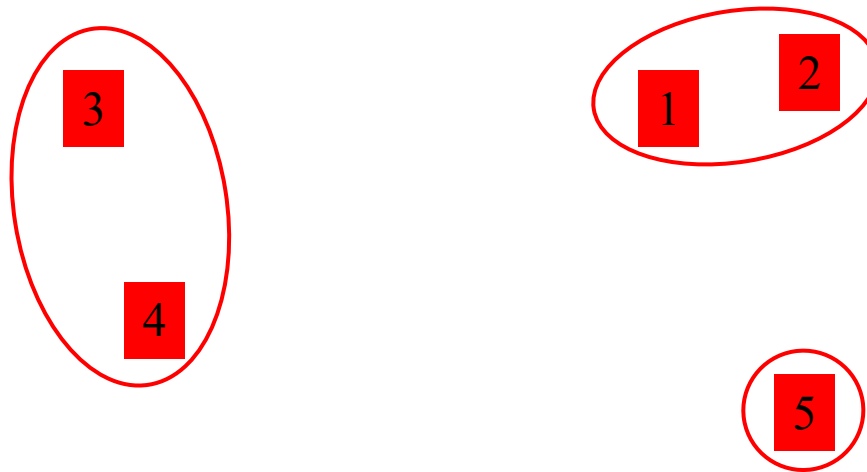
# Classification ascendante hiérarchique

- Illustration de l'agglomération progressive des éléments



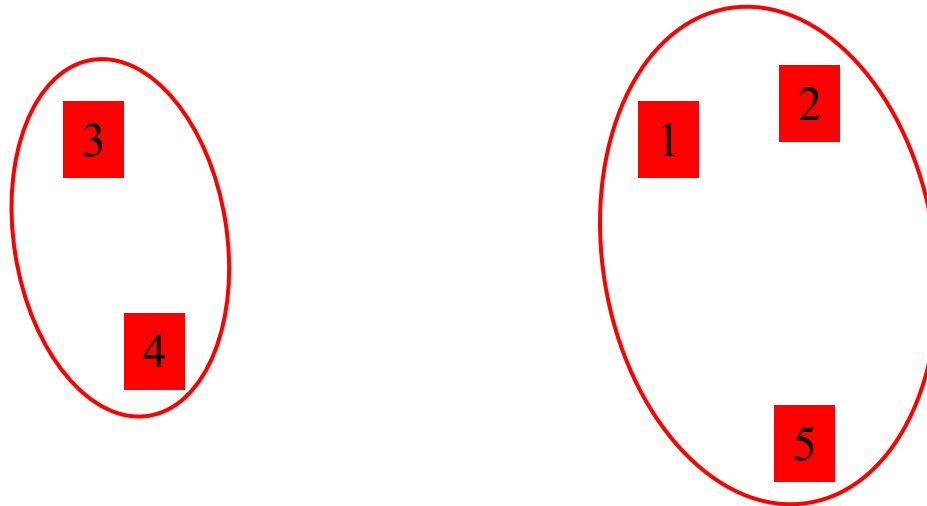
# Classification ascendante hiérarchique

- Illustration de l'agglomération progressive des éléments



# Classification ascendante hiérarchique

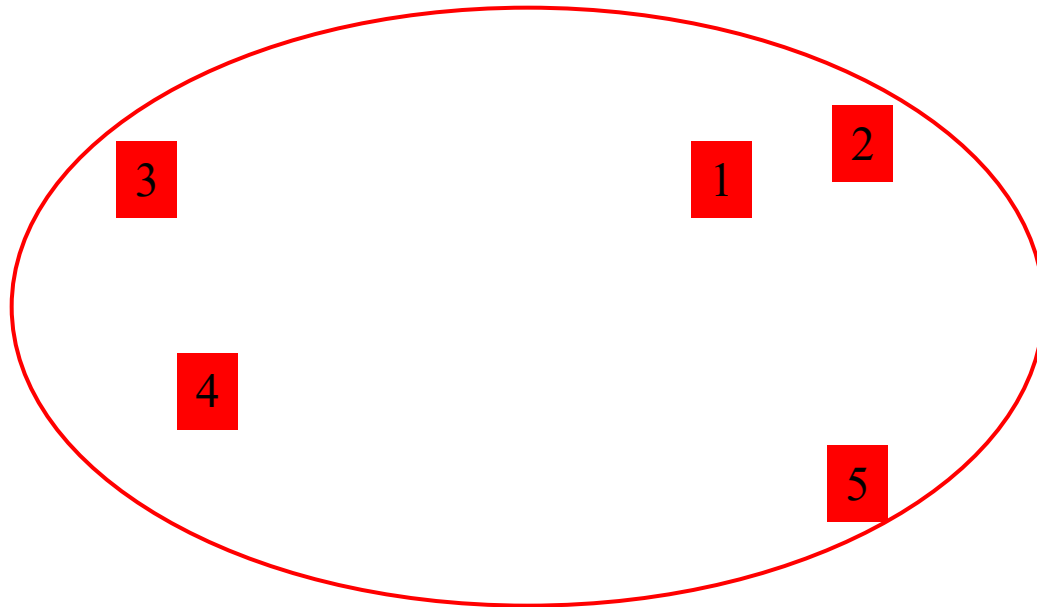
- Illustration de l'agglomération progressive des éléments





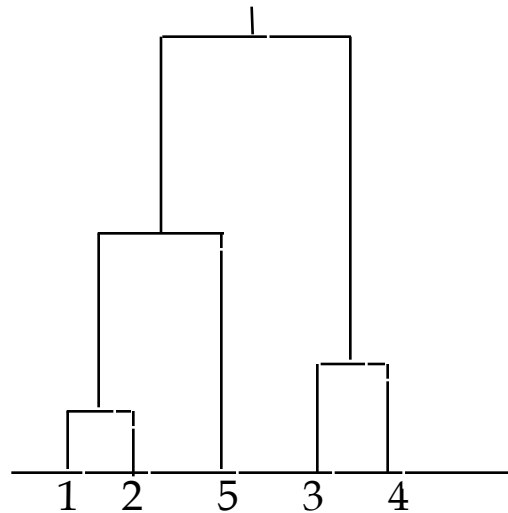
# Classification ascendante hiérarchique

- Illustration de l'agglomération progressive des éléments



# Classification ascendante hiérarchique

- Dendrogramme correspondant à l'agglomération illustrée précédemment

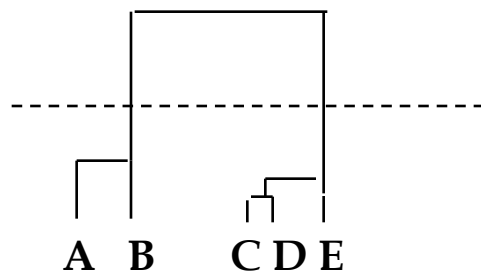


# Classification ascendante hiérarchique

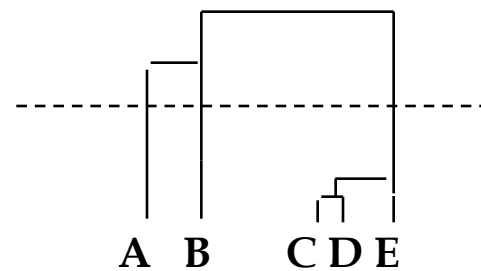
- Choix d'une partition à partir du dendrogramme
  - La CAH permet par étapes successives de créer la hiérarchie de partitions (dendrogramme)
  - Une partition est déterminée par coupure de l'arbre à un certain niveau d'agrégation

# Classification ascendante hiérarchique

- Choix d'une partition à partir du dendrogramme
  - Le niveau d'agrégation est important
    - Plus les classes sont agrégées avec un faible niveau, plus elles se ressemblent
    - Les partitions seront prises entre les classes dissemblables (agrégées avec un fort niveau)



Partition à 2 classes



Partition à 3 classes

# Classification ascendante hiérarchique

- Choix d'une partition à partir du dendrogramme
  - A toute partie  $h$  de la hiérarchie, une valeur numérique  $v(h)$  est associée telle que :
    - Si  $h \subset h'$  alors  $v(h) < v(h')$
  - Les valeurs généralement utilisées sont les distances correspondant à chaque étape d'agrégation
  - Ces valeurs sont appelées indices de niveau

# Classification ascendante hiérarchique

- Choix d'une partition à partir du dendrogramme
  - On représente généralement les indices de niveau par un diagramme en bâtons
  - Ce diagramme montre l'évolution des valeurs des indices lorsqu'on passe d'une partition donnée à une partition plus fine
  - L'allure de ce diagramme suggère des niveaux de coupure privilégiés
  - En effet, une augmentation importante de la valeur de l'indice de niveau suggère que la nouvelle partition est moins homogène que la partition de niveau inférieur

# Classifications non supervisées

- Introduction
- Terminologie
- Les centre mobiles
- Classification ascendante hiérarchique (CAH)
- ➔ ▪ Classification mixte : CAH + centres mobiles

# Classification mixte

- Les différentes méthodes de classification non supervisée offrent chacune des avantages et inconvénients
- La méthode des centres mobiles
  - Avantage : permet d'obtenir une partition sur un ensemble de données volumineux à un faible coût
  - Inconvénients :
    - La partition obtenue est dépendante des premiers centres choisis
    - Le nombre de classes doit être fixé a priori



# Classification mixte

- La méthode de classification ascendante hiérarchique
  - Avantages :
    - Méthode déterministe (donne toujours les mêmes résultats à partir des mêmes)
    - Donne des indications sur le nombre de classes à retenir
  - Inconvénient :
    - Elle est mal adaptée aux ensembles de données volumineux

# Classification mixte

- Il est donc intéressant de combiner ces deux approches pour obtenir une méthode plus performante
- La méthode des centres mobiles peut ainsi être utilisée comme auxiliaire de la méthode de classification ascendante hiérarchique

# Classification mixte

- Principe de l'algorithme
  - Il est composé de trois étapes :
    - Etape 1 : partitionnement initial (méthode des centres mobiles)
    - Etape 2 : agrégation hiérarchique des classes (CAH)
    - Etape 3 : optimisation de la partition obtenue (méthode des centres mobiles)

# Classification mixte

- Principe de l'algorithme
  - Etape 1 : partitionnement initial
    - Objectif : obtenir rapidement et à faible coût une partition des  $n$  objets en  $k$  classes homogènes
    - $k$  doit être plus élevée que le nombre de classes désirées et plus petit que  $n$
    - Principe : utilisation de la méthode des centres mobiles

# Classification mixte

- Principe de l'algorithme

- Etape 2 : agrégation hiérarchique des classes
  - Une classification ascendante hiérarchique est réalisée
  - Les éléments terminaux de l'arbre sont les  $k$  classes de la partition obtenue précédemment
  - Cette étape permet
    - De reconstituer les classes qui ont été fragmentées
    - D'agréger des éléments dispersés autour de leur centre de gravité
  - L'arbre est construit selon le critère de Ward

# Classification mixte

- Principe de l'algorithme
  - Etape 3 : optimisation de la partition obtenue
    - La partition finale est obtenue par coupure de l'arbre
    - L'homogénéité des classes obtenues est optimisée par réaffectation des objets par la méthode des centres mobiles

# Références

- G. Govaert, « Analyse des données », Ed Hermes Science, 2003
- L. Lebart, A. Morineau, M. Piron, « Statistique exploratoire multidimensionnelle », Ed Dunod, 2004
- A.K. Jain, R.C. Dubes, « Algorithms for clustering data », Prentice Hall Inc., 1988
- Librairie scipy pour la mise en œuvre des méthodes :
  - Centres mobiles :  
<https://docs.scipy.org/doc/scipy/reference/generated/scipy.cluster.vq.kmeans2.html>
  - CAH :  
<https://docs.scipy.org/doc/scipy/reference/cluster.hierarchy.html>



ÉCOLE  
**CENTRALE** LYON

36, avenue Guy de Collongue 69130 Écully - France  
+33 (0)4 72 18 60 00

[www.ec-lyon.fr](http://www.ec-lyon.fr)