Polytech Nantes

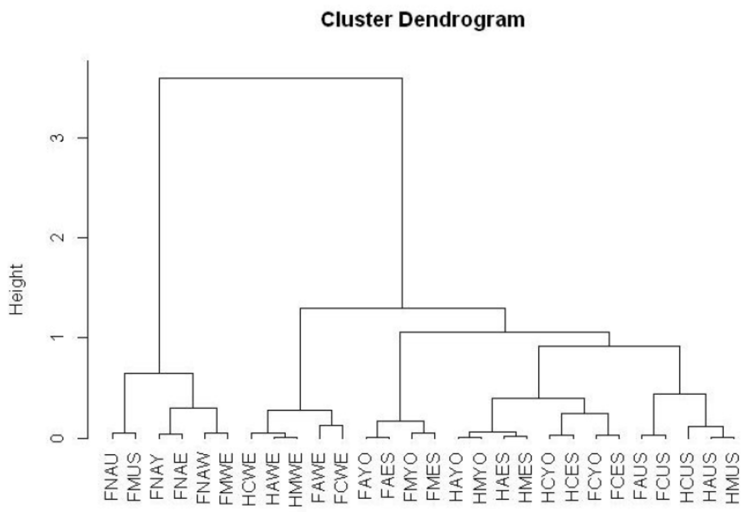
Analyse des données

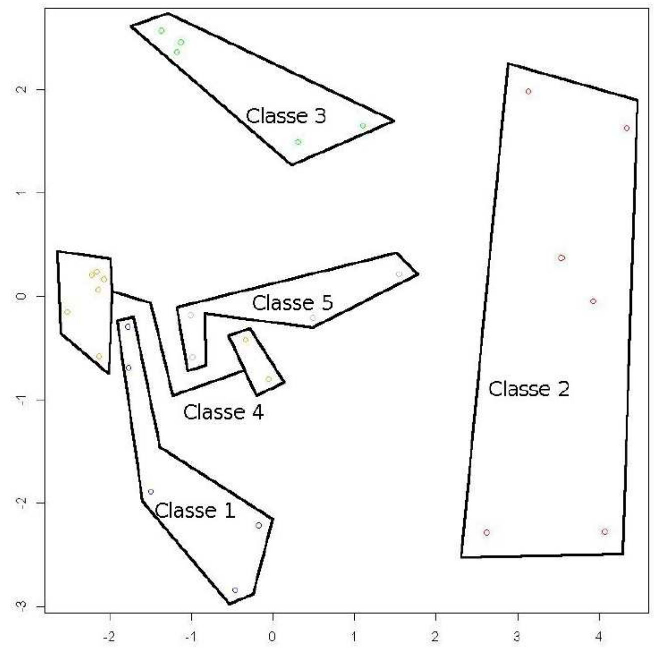
Classification 1

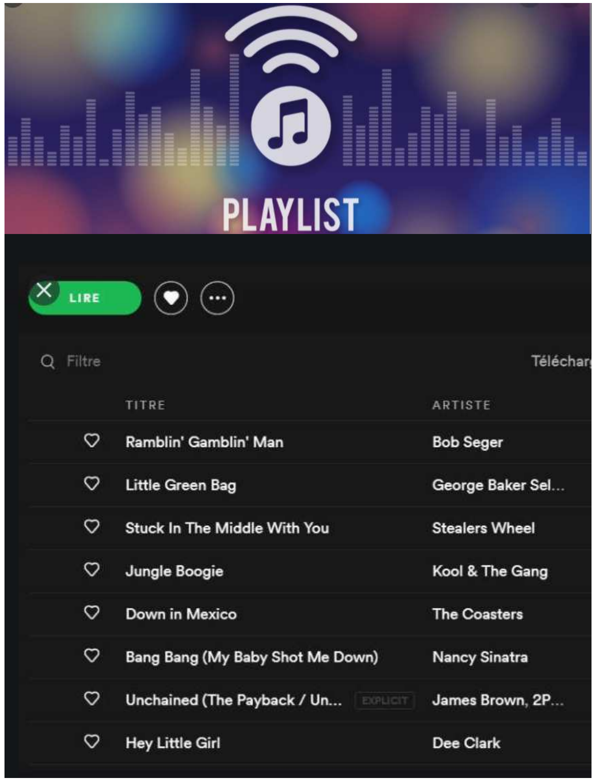
P. Kuntz

2021-2022

Analyse factorielle et classification







Question : comment classer votre playlist en un nombre « raisonnable » de classes ?

Classification de playlists

Classification de playlists

Objectifs de la classification

Georges Louis Leclerc, comte de Buffon (dans Histoire Naturelle, 1749)

« *La seule manière de construire une méthode instructive et naturelle est de*
- *mettre ensemble les choses qui sont similaires*
- *et séparées celles qui sont différentes* »

Conséquences

1. Comment mettre ensemble **et** séparer en même temps ?

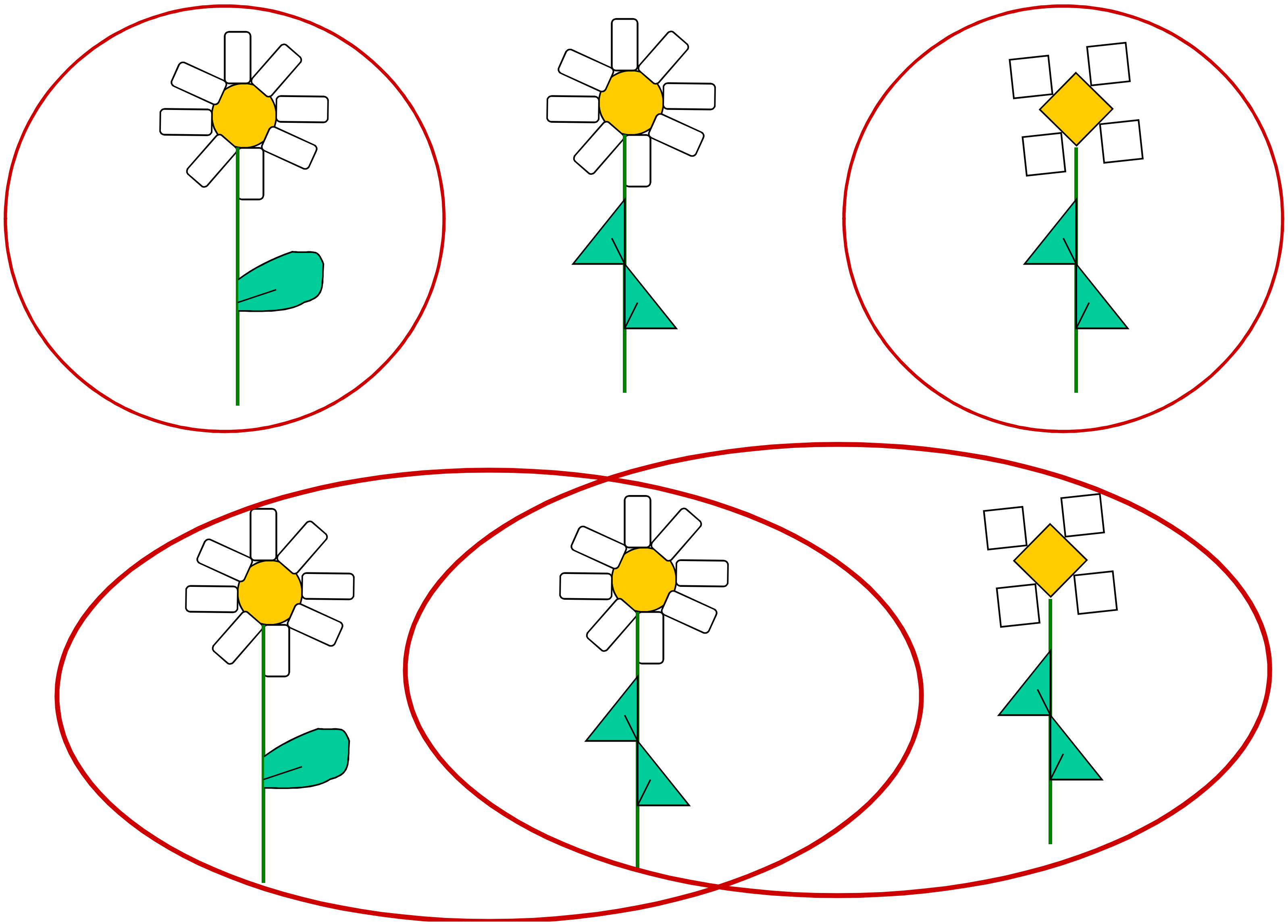
Principe de non-ambiguïté : l'intersection de deux classes n'est pas une troisième

2. Une classification dépend de deux critères : **homogénéité** à l'intérieur des classes, **séparation** des classes

Pas évident d'optimiser simultanément les deux critères

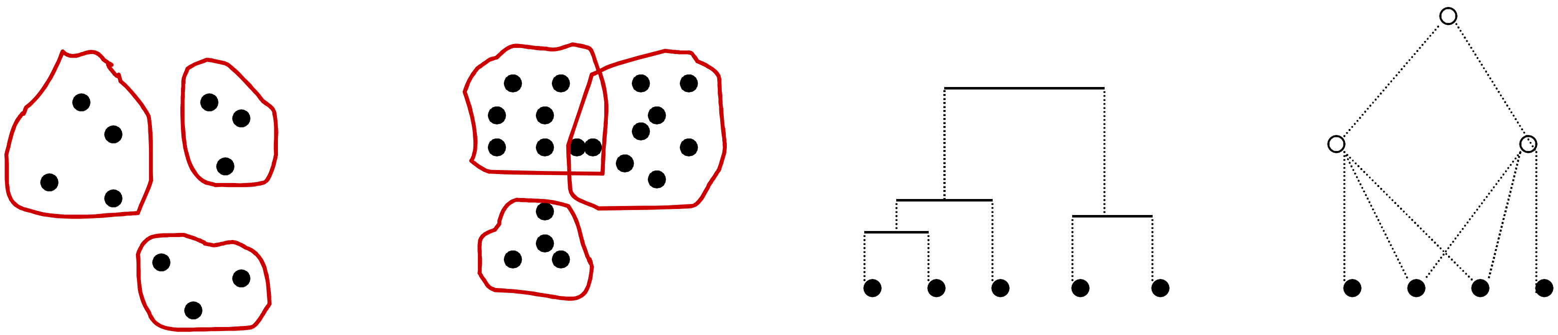
Principe de non-ambiguïté

Problème : hybridation (évolution)



Plan

- Modèles de classification



- Algorithmes de classification

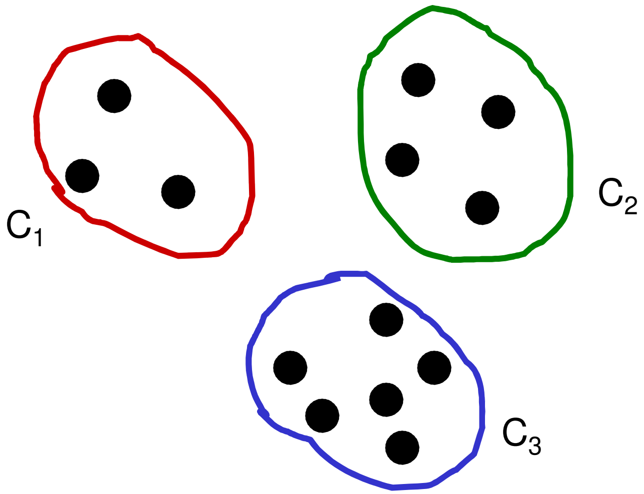
Modèles de classes I

<u>INPUT</u> : ensemble de données X

Partition de X : $P = \{C_1, C_2, \ldots, C_n\}$ telle que

i) $\cup_i C_i = X$

ii) $C_i \cap C_j = \varnothing$ pour tout C_i, C_j

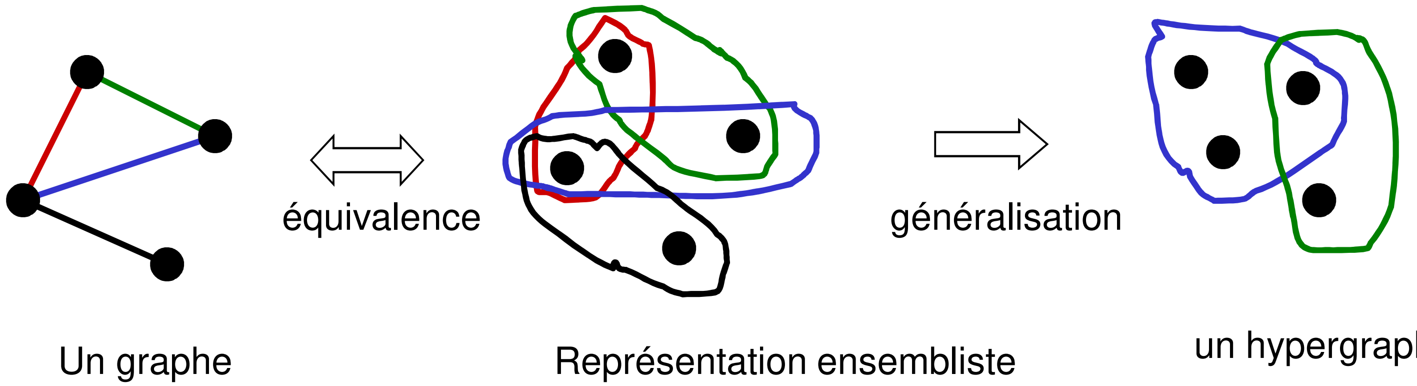


Modèles de classes II

<u>INPUT</u> : ensemble de données X

Hypergraphe : H = (X, E)

$E = \{X_1, X_2, \ldots, X_n\}$ parties de X ; $\cup_i X_i = X$



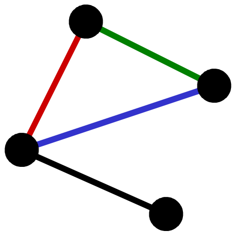
Graphe : G = (X, E)

un hypergraphe tel que si $A \in E$ alors card(A) = 2

Hypergraphe : $H = (X, E)$

$E = \{X_1, X_2, \dots, X_n\}$ parties de X ; $\cup_i X_i = X$

Graphe : $G = (X, E)$ E :



Un graphe

Modèles de classes III

Système de classes de X : K = (X, C) hypergraphe tel que.

éléments de C : classes

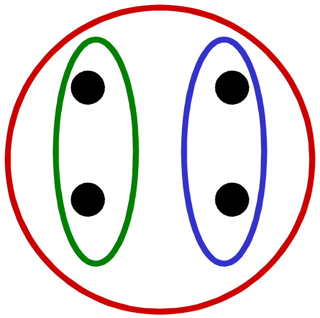
i) $\emptyset \notin$ C

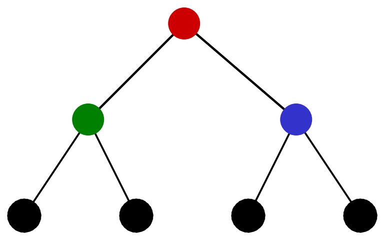
ii) X $\in$ C

iii) $\{x\} \in$ C for any x $\in$ X

Hiérarchie : Système de classes K qui satisfait la condition suivante

si A $\in$ K et B $\in$ K alors $A \cap B \in \{A, B, \emptyset\}$





Un arbre planté

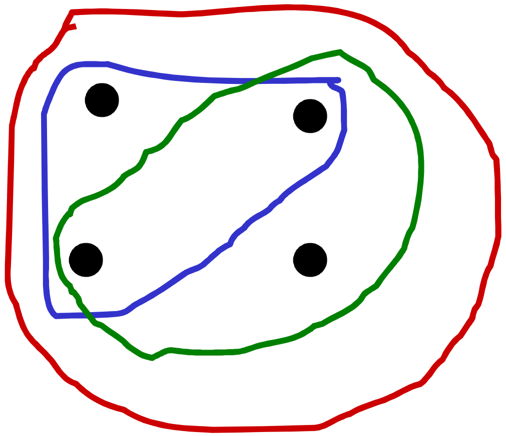
- aucun recouvrement (l'intersection de 2 classes ne peut pas être une 3ème classe)

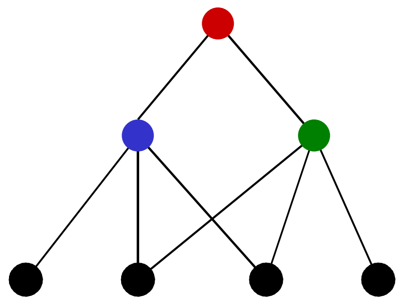
Modèles de classes IV

Hiérarchie faible (Bandelt & Dress, 1989)

Système de classes K qui satisfait la condition suivante

si A, B, C $\in$ K alors $A \cap B \cap C \in \{A \cap B, A \cap C, B \cap C\}$

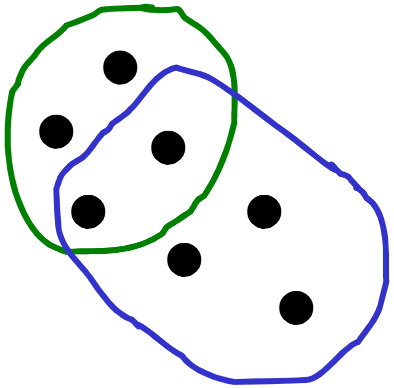




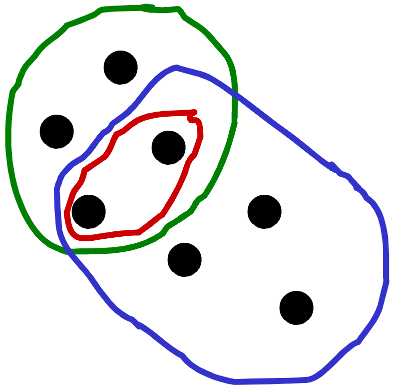
L'intersection de 3 classes est toujours l'intersection de 2 d'entre elles

Modèles de classes V

Hypergraphe fermé : Hypergraphe H = (X, E) tel que
pour tout A, B $\in$ E t.q. $A \cap B \neq \emptyset$ alors $A \cap B \in$ E



un hypergraphe

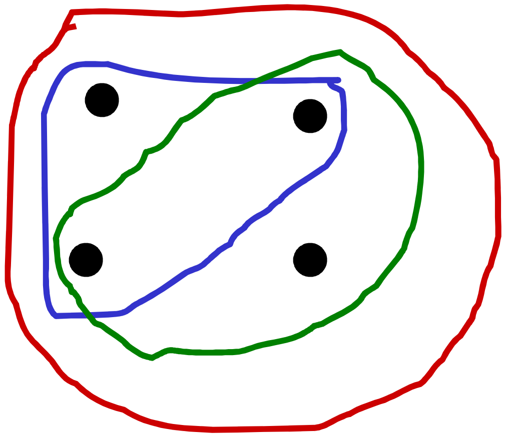


un hypergraphe fermé

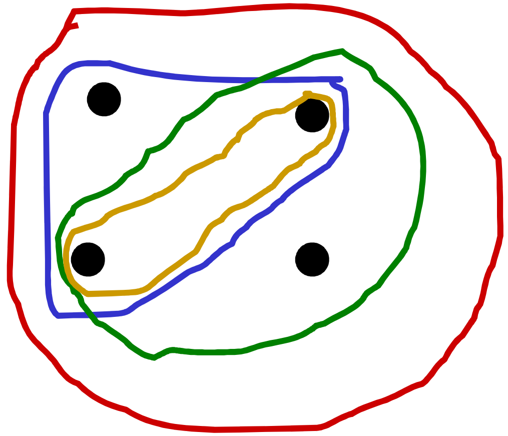
Modèles de classes VI

Quasi-hiérarchie (Diatta & Fichet, 1994)

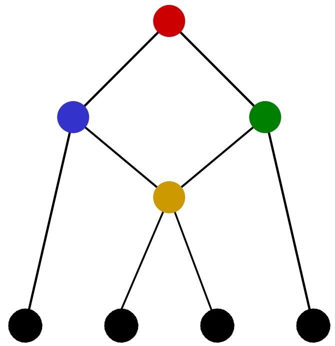
hiérarchie faible fermée



Une hiérarchie faible



Une quasi-hierarchie



Modèles de classes

- partitions
- hiérarchies

algorithmes

- algorithmes de réalocation itérative (K-means)
- Classification hiérarchique