

XAI

eXplainable Artificial Intelligence

Cours 9 - mardi 5 décembre 2023

Marie-Jeanne Lesot
Christophe Marsala
Jean-Noël Vittaut
Gauvain Bourgne

LIP6, Sorbonne Université

XAI non supervisée : expliquer les données

- Objectif global
 - comprendre le contenu d'un ensemble de données collectées
 - besoin de le **rendre intelligible**
- Principe : **résumer les données** et les caractériser
 - mettre en évidence leur structure sous-jacente
 - identifier de corrélation entre attributs
 - obtenir une représentation simplifiée
- Différentes approches
 - décomposition en sous-groupes homogènes et distincts
 - construction de résumés linguistiques
 - mise en évidence de cas particuliers

XAI non supervisée : expliquer les données

- Objectif global
 - comprendre le contenu d'un ensemble de données collectées
 - besoin de le **rendre intelligible**
- Principe : **résumer les données** et les caractériser
 - mettre en évidence leur structure sous-jacente
 - identifier de corrélation entre attributs
 - obtenir une représentation simplifiée
- Différentes approches
 - décomposition en sous-groupes homogènes et distincts : clustering
 - construction de résumés linguistiques : data-to-text
 - mise en évidence de cas particuliers : détection d'exceptions

XAI non supervisée : expliquer les données

- Objectif global
 - comprendre le contenu d'un ensemble de données collectées
 - besoin de le **rendre intelligible**
- Principe : **résumer les données** et les caractériser
 - mettre en évidence leur structure sous-jacente
 - identifier de corrélation entre attributs
 - obtenir une représentation simplifiée
- Différentes approches
 - décomposition en sous-groupes homogènes et distincts : **clustering**
 - construction de résumés linguistiques : data-to-text
 - mise en évidence de cas particuliers : détection d'exceptions

Clustering

- **Objectif** : décomposer les données en sous-groupes, **clusters**

Clustering

- **Objectif** : décomposer les données en sous-groupes, **clusters**
 - **homogènes** : même cluster \implies objets similaires
 - **distincts** : clusters différents \implies objets dissimilaires

Clustering

- **Objectif** : décomposer les données en sous-groupes, **clusters**
 - **homogènes** : même cluster \implies objets similaires
 - **distincts** : clusters différents \implies objets dissimilaires
- **Utilité**
 - mise en évidence de la structure cachée des données
 - représentation simplifiée des données : nombre réduit de groupes homogènes au lieu d'un nombre élevé d'objets

Clustering

- **Objectif** : décomposer les données en sous-groupes, **clusters**
 - **homogènes** : même cluster \implies objets similaires
 - **distincts** : clusters différents \implies objets dissimilaires
- **Utilité**
 - mise en évidence de la structure cachée des données
 - représentation simplifiée des données : nombre réduit de groupes homogènes au lieu d'un nombre élevé d'objets
- **Cadre XAI** : de plus, **descriptions intelligibles des clusters**
 - visualisation de la structure des clusters :
 - p. ex. dendrogramme du clustering hiérarchique
 - caractérisation par régions
 - approche par substitution, score d'importance d'attribut, subspace clustering
 - example-based explanation: caractérisation par représentant
 - construction de prototypes

Clustering

- Nombreux algorithmes !

Clustering

- Nombreux algorithmes !
- Approches par partitionnement : GMM, k-moyennes et ses variantes
 - principe : déterminer la “meilleure” partition en un nombre fixé de clusters
- Clustering spectral : diagonalisation de la matrice de similarité
 - principe : deux données se ressemblent si elles ont le même profil de similarité, i.e. si elles ressemblent aux mêmes données
- Clustering hiérarchique
 - principe : multiples partitions, de niveaux de granularité différents
- Approches par densité : DBSCAN
 - principe : estimation de la densité locale des données
- Deep clustering : architecture encodeur-décodeur
 - coût de reconstruction standard ou avec préservation de la localité
 - exploitation de l'espace latent, p.ex. k -moyennes

Au programme du jour

- 1. Clustering par partitionnement
- 2. Autres approches de clustering
- 3. Prototypes : explication de clusters par l'exemple
- 4. Exemples d'applications

Au programme du jour

- 1. Clustering par partitionnement
 - GMM
 - k -moyennes
 - autres variantes pondérées
- 2. Autres approches de clustering
- 3. Prototypes : explication de clusters par l'exemple
- 4. Exemples d'applications

GMM

- Estimation paramétrique de la densité de distribution des données

$$p(x) = \sum_{r=1}^k \pi_r \mathcal{N}(x|\mu_r, \Sigma_r)$$

$$\mathcal{N}(x|\mu_r, \Sigma_r) = \frac{1}{\sqrt{(2\pi)^k |\Sigma_r|}} \exp\left(-\frac{1}{2}(x - \mu_r)^T \Sigma_r^{-1} (x - \mu_r)\right)$$

$$\sum_{r=1}^k \pi_r = 1$$

- Fonction de coût : maximiser la **vraisemblance** des données

$$\ln p(X) = \sum_{i=1}^n \ln \left(\sum_{r=1}^k \pi_r \mathcal{N}(x_i|\mu_r, \Sigma_r) \right)$$

GMM

- Optimisation : **EM** Expectation-Maximisation

$$u_{ir} = \frac{\pi_r \mathcal{N}(x_i | \mu_r, \Sigma_r)}{\sum_{s=1}^k \pi_s \mathcal{N}(x_i | \mu_s, \Sigma_s)}$$

$$\mu_r = \frac{1}{n_r} \sum_{i=1}^n u_{ir} x_i$$

$$\Sigma_r = \frac{1}{n_r} \sum_{i=1}^n u_{ir} (x_i - \mu_r)(x_i - \mu_r)^T$$

$$\pi_r = \frac{n_r}{n} \quad \text{avec} \quad n_r = \sum_{i=1}^n u_{ir}$$

Algorithme des k -moyennes

- Par rapport aux GMM
 - affectation : pondérée \longrightarrow binaire
 - distance : Mahalanobis \longrightarrow euclidienne
 - fonction de coût : vraisemblance \longrightarrow quantification
- Notations
 - données : $x_i, i = 1 \dots n$
 - centres : $w_r, r = 1 \dots k$
 - affectation :

$$u_{ir} = \begin{cases} 1 & \text{si } x_i \in C_r \\ 0 & \text{sinon} \end{cases}$$

Algorithme des k -moyennes

- **Fonction de coût : erreur de quantification**
 - erreur commise quand on représente x_i par le centre du cluster auquel elle est affectée

$$J = \sum_{r=1}^k \sum_{i=1}^n u_{ir} \|x_i - w_r\|^2$$

Algorithme des k -moyennes

- **Fonction de coût : erreur de quantification**
 - erreur commise quand on représente x_i par le centre du cluster auquel elle est affectée

$$J = \sum_{r=1}^k \sum_{i=1}^n u_{ir} \|x_i - w_r\|^2$$

- Sous contraintes
 - chaque donnée appartient à un et un seul cluster
 - aucun cluster n'est vide

$$\forall i = 1 \dots n \quad \sum_{r=1}^k u_{ir} = 1$$

$$\forall r = 1 \dots k \quad \sum_{i=1}^n u_{ir} > 0$$

Optimisation alternée

$$J = \sum_{r=1}^k \sum_{i=1}^n u_{ir} \|x_i - w_r\|^2 \quad \text{sous contraintes}$$

- Mise à jour itérative
 - optimisation des u_{ir} pour w_r fixés

Optimisation alternée

$$J = \sum_{r=1}^k \sum_{i=1}^n u_{ir} \|x_i - w_r\|^2 \quad \text{sous contraintes}$$

- Mise à jour itérative
 - optimisation des u_{ir} pour w_r fixés

$$u_{ir} = \begin{cases} 1 & \text{si } r = \arg \min_{s=1\dots k} \|x_i - w_s\| \\ 0 & \text{sinon} \end{cases}$$

Optimisation alternée

$$J = \sum_{r=1}^k \sum_{i=1}^n u_{ir} \|x_i - w_r\|^2 \quad \text{sous contraintes}$$

- Mise à jour itérative
 - optimisation des u_{ir} pour w_r fixés

$$u_{ir} = \begin{cases} 1 & \text{si } r = \arg \min_{s=1\dots k} \|x_i - w_s\| \\ 0 & \text{sinon} \end{cases}$$

- optimisation des w_r pour u_{ir} fixés

Optimisation alternée

$$J = \sum_{r=1}^k \sum_{i=1}^n u_{ir} \|x_i - w_r\|^2 \quad \text{sous contraintes}$$

- Mise à jour itérative
 - optimisation des u_{ir} pour w_r fixés

$$u_{ir} = \begin{cases} 1 & \text{si } r = \arg \min_{s=1\dots k} \|x_i - w_s\| \\ 0 & \text{sinon} \end{cases}$$

- optimisation des w_r pour u_{ir} fixés

$$w_r = \frac{\sum_{i=1}^n u_{ir} x_i}{\sum_{i=1}^n u_{ir}}$$

k -moyennes

(Ball & Hall, 66)

- Paramètre
 - k le **nombre de clusters** recherchés
- Entrée
 - la matrice des données $x_i, i = 1...n$
- Initialisation : choix de k centres $w_r, r = 1...k$
- Itération : jusqu'à stabilisation des positions des centres
 - **mise à jour des affectations** :
chaque donnée est affectée au centre le plus proche
 - **mise à jour des centres** :
chaque centre = moyenne des données qui lui sont affectées

k -moyennes

(Ball & Hall, 66)

- Paramètre
 - k le **nombre de clusters** recherchés
- Entrée
 - la matrice des données $x_i, i = 1 \dots n$
- **Résultat**
 - les **centres** des clusters $w_r, r = 1 \dots k$
 - la **matrice d'affectation** qui indique, pour chaque donnée x_i et chaque cluster C_r , si $x_i \in C_r$

$$u_{ir} = \begin{cases} 1 & \text{si } x_i \in C_r \\ 0 & \text{sinon} \end{cases}$$

Quelques limites des k -moyennes

- Pas de garantie d'obtention de l'optimum global
 - obtention d'un minimum local
 - nécessité de lancer l'algorithme plusieurs fois, avec plusieurs initialisations
 - étude des propriétés de convergence (Bottou et Bengio, 95)
- Sensibilité aux exceptions (*outliers*)
 - erreur quadratique
- Recherche de clusters sphériques

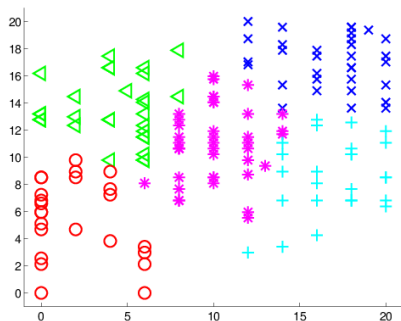
fc_m: fuzzy *c*-means

- Motivation : **traitement des cas ambigus**
 - points à égale distance de deux clusters
 - chevauchement de clusters
- Principe : cf cours 7 théorie des sous-ensembles flous
 - appartenance floue de x_i au cluster C_r
 - ⇒ un point peut appartenir à plusieurs clusters
 - représentée par un **degré d'appartenance** : $u_{ir} \in [0, 1]$
 - sémantique de multi-appartenance \neq appartenance inconnue
- Référence initiale : Bezdek 1973



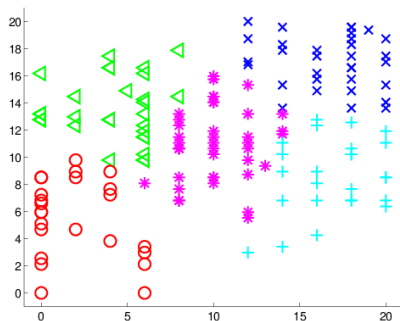
Exemple : notes d'étudiants à deux ue

● Partition classique

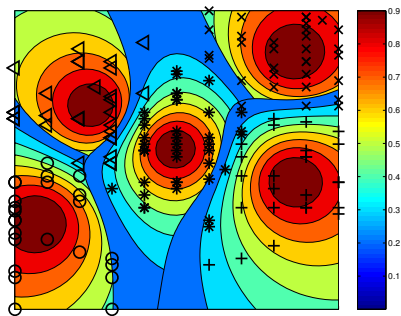


Exemple : notes d'étudiants à deux ue

● Partition classique



● Partition floue



Fonction de coût des c -moyennes floues

- Définie comme l'erreur de quantification

$$J = \sum_{r=1}^c \sum_{i=1}^n u_{ir}^m \|x_i - w_r\|^2 \quad \text{avec } u_{ir} \in [0, 1]$$

- $m = \text{fuzzifier}$: détermine le **degré de flou de la partition**
 - le plus souvent $m = 2$

- Sous contraintes

- aucun cluster n'est vide : $\forall r = 1 \dots c \quad \sum_{i=1}^n u_{ir} > 0$

- même poids de chaque donnée : $\forall i = 1 \dots n \quad \sum_{r=1}^c u_{ir} = 1$

Optimisation de la fonction de coût

$$J = \sum_{r=1}^c \sum_{i=1}^n u_{ir}^m \|x_i - w_r\|^2 \text{ avec } u_{ir} \in [0, 1]$$

$$\text{sous contrainte } \forall i = 1 \dots n \quad \sum_{r=1}^c u_{ir} = 1$$

- Exercice : établir les formules de mises à jour de u_{ir} et de w_r

Optimisation des c -moyennes floues

- Notation $d_{ir} = \|x_i - w_r\|$
- **Optimisation alternée** : mise à jour itérative
 - dans le cas de la distance euclidienne
- optimisation des u_{ir} pour w_r fixés
- optimisation des w_r pour u_{ir} fixés

$$u_{ir} = \frac{1}{\sum_{s=1}^c \left(\frac{d_{ir}}{d_{is}} \right)^{\frac{2}{m-1}}}$$

$$w_r = \frac{\sum_{i=1}^n u_{ir}^m x_i}{\sum_{i=1}^n u_{ir}^m}$$

- Si $\exists r$ tel que $d_{ir} = 0$,
 - poser $u_{ir=1}$ et $u_{is} = 0$ pour tout $s \neq r$

L'algorithme des c -moyennes floues

- Paramètres
 - c le nombre de clusters recherchés
 - m le fuzzifier
- Entrée
 - la matrice des données $x_i, i = 1 \dots n$
- Initialisation : choix de c centres $w_r, r = 1 \dots c$
- **Itération** : jusqu'à stabilisation des positions des centres
 - **mise à jour de la matrice d'affectation**
 - **mise à jour des centres**
centre = moyenne pondérée des données
avec pour poids les degrés d'appartenance

L'algorithme des c -moyennes floues

- Paramètres
 - c le nombre de clusters recherchés
 - m le fuzzifier
- Entrée
 - la matrice des données
- Résultat
 - les centres des clusters
 - la **matrice d'affectation** qui indique, pour chaque donnée x_i et chaque cluster C_r , le **degré d'appartenance** de x_i à C_r

Caractéristiques des c -moyennes floues

- Définition **relative** des degrés d'appartenance

$$u_{ir} = \frac{1}{\sum_{s=1}^c \left(\frac{d_{ir}}{d_{is}} \right)^{\frac{2}{m-1}}}$$

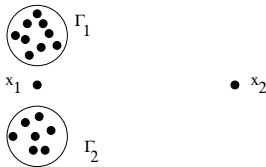
- dépend de la distance au cluster r
- mais aussi de la distance aux autres clusters
- **Propriétés**
 - méthode plus stable et plus robuste que les k -moyennes
 - moins sensible à l'initialisation que les k -moyennes
 - moins de risque de tomber dans un minimum local

c -moyennes possibilistes



(Krishnapuram & Keller, 93)

- **Motivation** : le problème



- degré d'appartenance : degré de **partage** entre les clusters
- dû à la normalisation des degrés d'appartenance

$$\forall i = 1 \dots n, \quad \sum_{r=1}^k u_{ir} = 1$$

c -moyennes possibilistes

- Fonction de coût

$$J = \sum_{r=1}^c \sum_{i=1}^n u_{ir}^m d_{ir}^2 + \sum_{r=1}^c \eta_r^2 \sum_{i=1}^n (1 - u_{ir})^m$$

$$\text{sous contrainte } \sum_{i=1}^n u_{ir} > 0$$

- nouveau terme : récompenser les degrés d'appartenance élevés, et éviter la solution triviale

- Equation de mise à jour** : exercice

c -moyennes possibilistes

- Fonction de coût

$$J = \sum_{r=1}^c \sum_{i=1}^n u_{ir}^m d_{ir}^2 + \sum_{r=1}^c \eta_r^2 \sum_{i=1}^n (1 - u_{ir})^m$$

sous contrainte $\sum_{i=1}^n u_{ir} > 0$

- nouveau terme : récompenser les degrés d'appartenance élevés, et éviter la solution triviale

- Equation de mise à jour**

$$u_{ir} = \frac{1}{1 + \left(\frac{d_{ir}}{\eta_r} \right)^{\frac{2}{m-1}}}$$

- pour $m = 2$, η_r = distance à partir de laquelle $u_{ir} < 0.5$

Comparaison des degrés des fc_m et pcm

$$\begin{array}{c} c\text{-moyennes floues} \\ u_{ir} = \frac{1}{\sum_{s=1}^c \left(\frac{d_{ir}}{d_{is}} \right)^{\frac{2}{m-1}}} \end{array}$$

$$\begin{array}{c} c\text{-moyennes possibilistes} \\ u_{ir} = \frac{1}{1 + \left(\frac{d_{ir}}{\eta_r} \right)^{\frac{2}{m-1}}} \end{array}$$

Cas **flou**

- degré de **partage** \implies effet de **répulsion** entre les clusters
 - si un cluster prend une partie des poids, il en reste moins pour les autres
 - désavantageux de partager les données
- problèmes dans le cas de données bruitées ou d'exceptions
 - distorsion des centres, attirés par les exceptions

Comparaison des degrés des *fc*m et *pc*m

$$c\text{-moyennes floues}$$

$$u_{ir} = \frac{1}{\sum_{s=1}^c \left(\frac{d_{ir}}{d_{is}} \right)^{\frac{2}{m-1}}}$$

$$c\text{-moyennes possibilistes}$$

$$u_{ir} = \frac{1}{1 + \left(\frac{d_{ir}}{\eta_r} \right)^{\frac{2}{m-1}}}$$

Cas **possibiliste**

- degré de **ressemblance**
- pas d'interaction entre les clusters : **définis indépendamment**
 - contributions à la fonction de coût indépendantes
- clusters potentiellement confondus
 - en fait, minimum global de J = tous les clusters confondus !
 - **c : nombre maximal de clusters souhaités**

Caractéristiques des *pcm*

- Peuvent manquer certaines zones
- Peuvent identifier des exceptions (*outliers*) :
 - ajustement individuel des poids
- **Mode-seeking** plus qu'algorithme de partitionnement :
identification de
 - clusters pertinents et représentatifs
 - et non d'une partition de toutes les données
- Sensibilité à l'initialisation
 - réalisée par les c -moyennes floues
 - les fcm donnent également des valeurs pour η_r

Variantes des pcm

- Combinaison avec les c -moyennes floues (Pal et al. 97; Pal et al. 04)
 - bénéficier des sémantiques de chacun des types de coefficients

$$J = \sum_{r=1}^c \sum_{i=1}^n (au_{ir}^m + bt_{ir}^\eta) d_{ir}^2 + \sum_{r=1}^c \eta_r^2 \sum_{i=1}^n (1 - t_{ir})^\eta$$

- avec la contrainte de sommation à 1 pour les u_{ir}

- Prise en compte d'un coefficient de répulsion entre les clusters (Timm et al, 04)

$$J = \sum_{r=1}^c \sum_{i=1}^n u_{ir}^m d_{ir}^2 + \sum_{r=1}^c \eta_r^2 \sum_{i=1}^n (1 - u_{ir})^m + \sum_{r=1}^c \gamma_r \sum_{s=1, s \neq r}^c \frac{1}{\xi d(w_r, w_s)^2}$$

Variantes des c -moyennes floues

- Autre contrainte de normalisation de la partition
 - c -moyennes possibilistes
- Autres fonctions de distance
 - pour détecter des formes de clusters spécifiques
 - pour traiter des données non vectorielles
- Autres fonctions de coût
 - pour prendre en compte des contraintes particulières
 - concernant le bruit ou les exceptions par exemple
- Estimation alternée de clusters
 - généralisation de l'optimisation alternée
 - pas de fonctions de coût
 - flexibilité accrue

Autres fonctions de distances

- Problématique
 - c -moyennes floues : clusters sphériques
 - changer la mesure de distance
 - donc changer la mise à jour des centres des clusters
- Algorithme de **Gustafson-Kessel** : distance de Mahalanobis

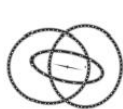
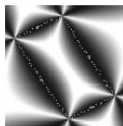
$$d^2(x_i, C_r) = {}^t(x_i - w_r)\Sigma_r^{-1}(x_i - w_r)$$

- clusters ellipsoïdaux, et non sphériques
- un cluster = un centre + une matrice de covariance
- équations de mise à jour pour w_r et Σ_r

Autres fonctions de distances

fuzzy shell clustering

- Principe
 - détecter des **structures plus complexes**
 - comme des lignes, des cercles, des ellipses
- Variantes
 - fuzzy c -varieties : clusters affines
un cluster = un point + ensemble de vecteurs unitaires orthogonaux
 - fuzzy c -quadric shell
 - fuzzy c -rectangles



(Höppner et al. 00)

Autres fonctions de distances

- Traitement de **données non vectorielles**
 - séquences, arbres, graphes
 - pas de centres des clusters : comment définir une moyenne (pondérée)
- Variantes **relationnelles** (Hathaway & Bezdek, 94)
 - entrée : matrice de distance
- **Variante à noyau** (Wu et al. 03)
 - entrée : matrice des produits scalaires
 - changement de représentation **implicite**

Autres fonctions de coût : traitement du bruit

- c -moyennes possibilistes (Krishnapuram & Keller, 93)
- Noise clustering (Davé, 91) : ajout d'un **cluster bruit**

- relâchement de la contrainte de sommation à 1

$$J = \sum_{r=1}^c \sum_{i=1}^n u_{ir}^m d_{ir}^2 + \sum_{i=1}^n \delta^2 \left(1 - \sum_{r=1}^c u_{ir} \right)^m$$

- **Estimateurs robustes** (Frigui et Krishnapuram, 96) :
 - remplacer la fonction des moindres carrés
 - ρ minimale en 0 et qui tend vers 0 pour les grandes valeurs

$$J = \sum_{r=1}^c \sum_{i=1}^n u_{ir}^m \rho_r(d_{ir})$$

- Introduction explicite de **poids de représentativité** (Keller, 00)

Autres fonctions de coût : changement de fuzzifier

$$J = \sum_{r=1}^c \sum_{j=1}^n u_{ir}^m d_{ir}^2$$

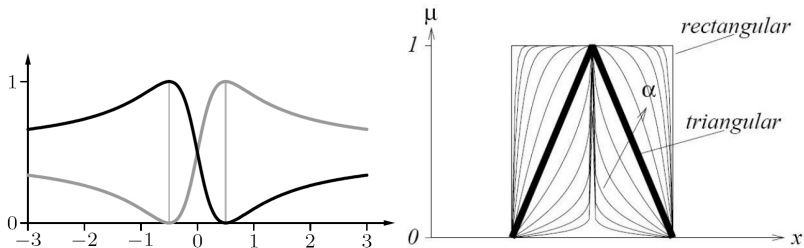
- Fuzzifier $m > 1 \Rightarrow u_{ir} \in [0, 1]$, mais en fait, $u_{ir} \in]0, 1[$
 - noyau : singleton, le centre
 - support : non borné
 - **toutes** les données influencent tous les centres
- Fonction g qui favorise les degrés extrêmes (Klawonn et Höppner, 03)

$$J = \sum_{r=1}^c \sum_{j=1}^n g(u_{ir}) d_{ir}^2$$

- par exemple $g(u_{ir}) = \alpha u_{ir} + (1 - \alpha) u_{ir}^2$ (Rousseeuw et al, 95)

Estimation alternée de clusters

- **Heuristiques** de mises à jour des centres ou des degrés d'appartenance
 - et non dérivée d'une fonction de coût
- Exemple : pour forcer la convexité



(Höppner et al. 00)

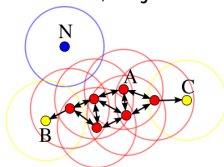
Au programme du jour

- 1. Clustering par partitionnement
- 2. **Autres approches de clustering**
 - DBSCAN : estimation non paramétrique de densité
 - clustering hiérarchique
- 3. Prototypes explications de clusters par l'exemple
- 4. Exemples d'applications

DBSCAN

(Ester et al. 96)

- Algorithme principal
 - tirer aléatoirement un point non encore visité x , et le marquer comme visité
 - construire son voisinage $\mathcal{N}_\epsilon(x) = \{z | d(x, z) \leq \epsilon\}$
 - si $|\mathcal{N}_\epsilon(x)| < minPts$, marquer x comme bruit
 - sinon, étendre le cluster($x, \mathcal{N}_\epsilon(x)$)
- Extension du cluster(x, \mathcal{N}) : pour chaque $z \in \mathcal{N}$
 - si z n'a pas été visité
 - marquer z comme visité
 - si $|\mathcal{N}_\epsilon(z)| \geq minPts$, $\mathcal{N} = \mathcal{N} \cup \mathcal{N}_\epsilon(z)$
 - si z n'est membre d'aucun cluster, l'ajouter au cluster en cours



source : wikipedia

DBSCAN

- Typologie des données
 - points centraux (core) : points dont le voisinage est dense
 - points frontières (border) : point non central dans le voisinage d'un point central
 - points aberrants (noise) : point non central et non frontière
- Paramètres : densité minimale pour la définition d'un cluster
 - ϵ taille du voisinage
 - *minPts* nombre de points minimum
- Variante : OPTICS

$$\text{coreDist}(x) = \begin{cases} \text{indéfini} & \text{si } |\mathcal{N}_\epsilon(z)| < \text{minPts} \\ \min(d(x, z)) & \text{sinon} \end{cases}$$

$$\text{reachabilityDist}(x, z) = \begin{cases} \text{indéfini} & \text{si } |\mathcal{N}_\epsilon(z)| < \text{minPts} \\ \max(\text{coreDist}(x), d(x, z)) & \text{sinon} \end{cases}$$

Clustering hiérarchique

- Approche agglomérative
 - initialiser un cluster par donnée
 - itérativement regrouper les clusters les plus proches

- Stratégies de chaînage

- minimale

$$d(C_1, C_2) = \min(d(x, z), x \in C_1, z \in C_2)$$

- complète

$$d(C_1, C_2) = \max(d(x, z), x \in C_1, z \in C_2)$$

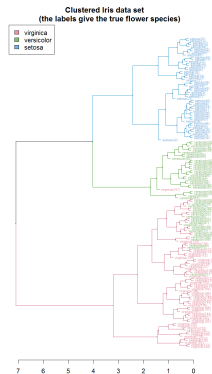
- moyenne

$$d(C_1, C_2) = \text{avg}(d(x, z), x \in C_1, z \in C_2)$$

- Ward

$$d(C_1, C_2) = \frac{n_1 n_2}{n_1 + n_2} d(\bar{x}_1, \bar{x}_2)$$

- Visualisation du dendrogramme



source : wikipedia

Au programme du jour

- 1. Clustering par partitionnement
- 2. Autres approches de clustering
- **3. Prototypes : explications de clusters par l'exemple**
- 4. Exemples d'applications

Résumé et interprétation

- **Caractériser** un ensemble de données
 - typiquement un cluster
 - aider à l'interprétation du résultat
- Description par une valeur
 - moyenne (pondérée), médiane, ...
 - problème : une valeur unique représente mal un ensemble
- Description enrichie
 - distribution : moyenne \rightarrow loi normale
 - problème : décrit le cluster dans son ensemble, n'est pas une caractérisation ou un résumé

Notion cognitive de prototype

(Rosch, 78)

- Un **prototype**
 - représente un groupe de données, le résume et le caractérise
 - capture à la fois
 - **les points communs** des membres du groupe
 - **leurs traits distinctifs** par rapport à d'autres groupes
- Notion de **typicalité**
 - toutes les données ne sont pas également représentatives
 - un individu typique
 - ressemble aux autres membres du groupe : **ressemblance interne**
 - diffère des membres des autres groupes : **dissimilarité externe**

Notion cognitive de prototype

(Rosch, 78)

- Un **prototype**
 - représente un groupe de données, le résume et le caractérise
 - capture à la fois
 - **les points communs** des membres du groupe
 - **leurs traits distinctifs** par rapport à d'autres groupes
- Notion de **typicalité**
 - toutes les données ne sont pas également représentatives
 - un individu typique
 - ressemble aux autres membres du groupe : **ressemblance interne**
 - diffère des membres des autres groupes : **dissimilarité externe**
- Contre-exemples pour les mammifères



Mise en œuvre

(Rifqi, 96)

Etant donné une catégorie C et un point $x \in C$

- **Ressemblance interne** R_C et **dissimilarité externe** D_C
 - r mesure de ressemblance, d mesure de dissimilarité

$$R_C(x) = \text{moy}(r(x, y), y \in C)$$

$$D_C(x) = \text{moy}(d(x, y), y \notin C)$$

- **Degré de typicalité** : agrégation de R_C et D_C

$$T_C(x) = \text{agr}(R_C(x), D_C(x))$$

- **Prototype** de C : agrégation des points les plus typiques

Mesures de ressemblance et de dissimilarité

- **Comparer la position relative de deux points**

- formellement : $r : \mathbb{R}^p \times \mathbb{R}^p \rightarrow [0, 1]$ $d : \mathbb{R}^p \times \mathbb{R}^p \rightarrow [0, 1]$
- définies en fonction d'une distance ou d'un produit scalaire
- non nécessairement duales

- Exemples :

$$d_1(x, y) = \frac{1}{\gamma_D} d(x, y) \qquad r_1(x, y) = 1 - \frac{1}{\gamma_R} d(x, y)$$

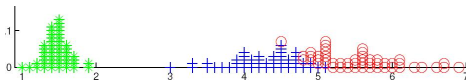
$$d_2(x, y) = 1 - \frac{1}{1 + \left(\frac{d(x, y)}{\gamma_D} \right)^\eta} \qquad r_2(x, y) = \frac{1}{1 + \left(\frac{d(x, y)}{\gamma_R} \right)^\eta}$$

Exemple des iris

- 150 fleurs, 4 attributs, 3 classes
 - longueur et largeur des pétales et sépales
 - 3 sortes d'iris *setosa*, *virginica*, *versicolor*



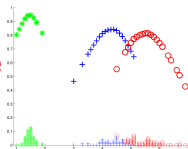
- On considère un seul attribut



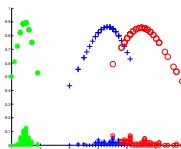
Comparaisons

Ressemblance
interne

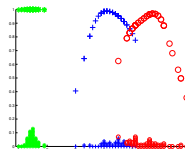
Euclidienne



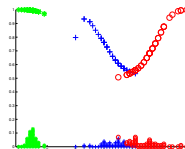
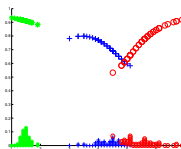
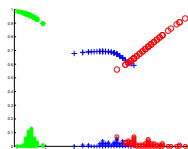
Cauchy



Fermi-Dirac



Dissimilarité
externe



Opérateurs d'agrégation

typicalité = *agr*(ressemblance interne, dissimilarité externe)

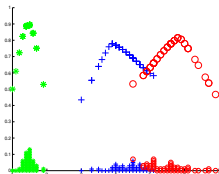
- Opérateurs conjonctifs : attitude sévère
 - t-normes, p. ex. $\min(R, D)$, $R \cdot D$, $\max(R + D - 1, 0)$
- Opérateurs disjonctifs : double sémantique
 - p. ex t-conormes, $\max(R, D)$, $R + D - R \cdot D$, $\min(R + D, 1)$
- Opérateurs de compromis : propriété de compensation
 - moyennes arithmétique, géométrique, pondérée
 - Ordered Weighted Average OWA (Yager, 88)
- Opérateurs à attitude variable : propriété de renforcement
 - exemple mica $\max(0, \min(1, k + (R - k) + (D - k)))$
 - somme symétrique $g(R, D)/(g(R, D) + g(1 - R, 1 - D))$
avec $g(x, y) = xy$

(Eude, 98; Detyniecki, 00)

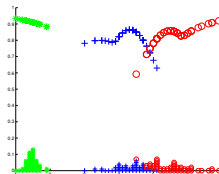
Application aux iris

typicalité = $\text{agr}(\text{ressemblance interne}, \text{dissimilarité externe})$

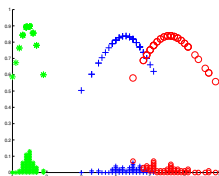
Conjonction
e.g. min,
sévère



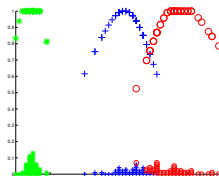
Disjonction
e.g. max,
double
sémantique



Compromis
e.g.
moyenne
pondérée,
propriété de
compensation



Attitude
variable
e.g. *mica*,
propriété de
renforcement

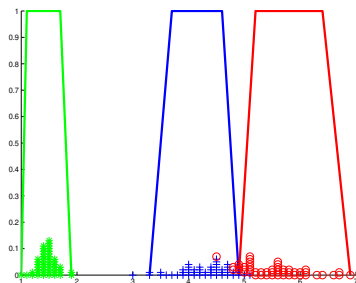


Agrégation en un prototype

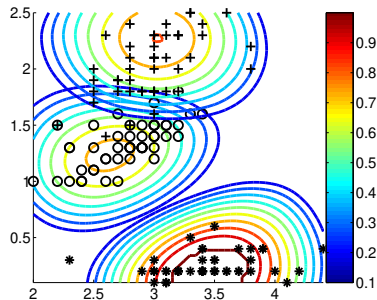
- **Prototype = agrégation des données les plus typiques**
 - globalement ou attribut par attribut
 - comme l'individu (la valeur) qui maximise la typicalité
 - par agrégation des individus (valeurs), pondérés par leurs degrés de typicalité
- **Prototype = notion imprécise**
 - exemple: *l'étudiant typique mesure environ 1m75*
 - mieux modélisée par un **sous-ensemble flou**
- Agrégation des données les plus typiques en un ensemble flou :
 - noyau = points de degré de typicalité > 0.9
 - support = points de degré de typicalité > 0.7
 - interpolation linéaire dans la partie intermédiaire

Applications aux iris

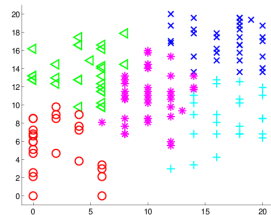
En 1D



En 2D



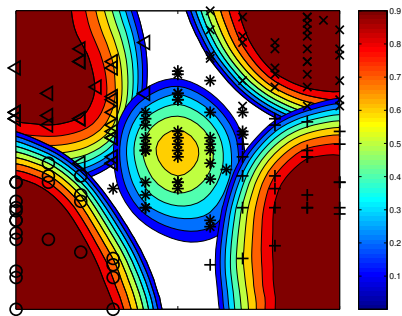
Autre exemple : notes



- Mesure de comparaison
 - dissimilarité : $d(x, y) = \min \left(1, \frac{\|x - y\|}{Z_d} \right)$
avec Z_d = demi-diamètre des données
 - ressemblance : $r(x, y) = \max \left(0, 1 - \frac{\|x - y\|}{Z_r} \right)$
avec Z_r = diamètre maximal des clusters
- Opérateur d'agrégation :
 - opérateur à attitude variable : somme symétrique

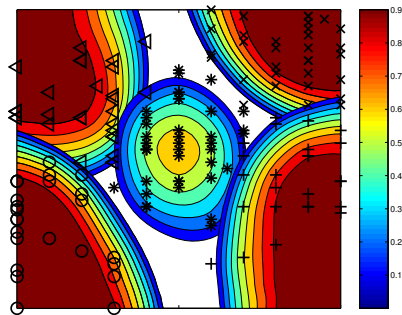
Résultat

prototypes flous

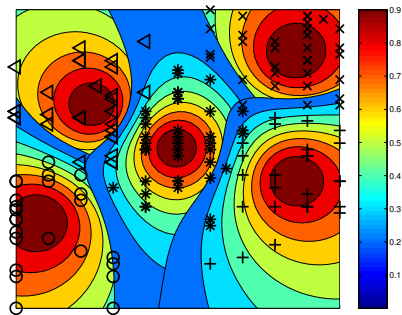


Résultat

prototypes flous



c -moyennes floues



Au programme du jour

- 1. Clustering par partitionnement
- 2. Autres approches de clustering
- 3. Prototypes explications de clusters par l'exemple
- 4. Exemple d'applications

Design émotionnel

- Importance de la forme dans le **design émotionnel**
 - rôle majeur dans la perception des objets créés
 - étape essentielle de la phase générative du design



Design émotionnel

- Description linguistique subjective : 40 adjectifs émotionnels

abstrait, acidulé, aérien, agressif, apaisant, attrayant, austère, chaud, classique, de qualité, doux, dynamique, élégant, équilibré, extravagant, féminin, fluide, libre, futuriste, géométrique, glamour, high-tech, jeune, ludique, luxueux, masculin, mystérieux, naturel, original, poétique, puissant, pur, rafraîchissant, rétro, séduisant, sécurisant, sensuel, simple, sophistiqué, sport

Designers	1	2	3	4	5	6	7	8	9
1. abstrait	X								
2. acidulé				X	X				
3. aérien			X						
4. agressif					X				
5. apaisant							X		X
6. attrayant									
7. austère		X				X			
8. chaud					X				
9. classique	X		X	X		X			

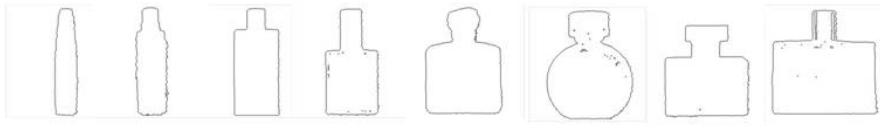
- Description numérique objective : descripteurs de forme globale

largeur, hauteur, périmètre, aire, circularité, compacité, élongation, moments invariants

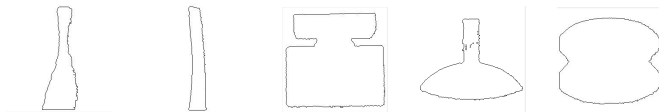


Design émotionnel

- Prototypes de forme
 - **8 formes principales**, principalement définies en termes d'élongation



- 17 exceptions, par exemple






Design émotionnel

- Caractérisation sémantique des types de formes
 - identification de **termes significativement présents ou absents**
 - critère numérique : équivalent du *lift*

$$\frac{\text{fréquence du terme } t \text{ dans le cluster } c}{\text{fréquence globale du terme } t}$$

Design émotionnel

Prototype	Présence significative	Absence significative
	éthéré acidulé dynamique mystérieux jeune sport	puissant attrayant extravagant agressif high-tech
	sport high-tech	glamour séduisant chaud
	puissant	fluide sensuel mystérieux ludique

Parmi les questions non discutées

- Choix des paramètres
- Mesures de qualité et validation
- Gain en intelligibilité ?