Apprentissage de Fonctions de Croyance à partir de Dissimilarités

Thierry Denœux

Université de Technologie de Compiègne
HEUDIASYC (UMR CNRS 6599)
http://www.hds.utc.fr/~tdenoeux

Problèmes posés

- Similarité/dissimilarité, distance : notions premières en perception, reconnaissance de formes, apprentissage
- Soient :
  - n objets $o_1,\ldots,o_n$
  - Une matrice de dissimilarités $D=(d_{ij})$, $d_{ij} \in \Delta$ entre ces n objets
  - d_{ij} = dissimilarité entre les objets i et j
- Les dissimilarités (distances) peuvent être calculées à partir d'attributs décrivant chacun des objets, ou être observées directement (analyse sensorielle par exemple).
- Propriétés :
  - Symétrie : $d_{ij}=d_{ji}$
  - (Inégalité triangulaire : Si $d_{ik} \leq d_{ij}+d_{jk} \rightarrow$ matrice de distances)

Problèmes posés (suite)

- Les n objets sont répartis en K classes (catégories) $\Omega=\{\omega_1,\ldots,\omega_K\}$
- Deux cas :
  - On dispose d'une information concernant la classe des n objets de la base d'apprentissage (issus d'une population), et on cherche à en déduire une règle de classement de nouveaux objets issus de la même population $\rightarrow$ apprentissage complètement ou partiellement supervisé
  - On ne dispose pas d'information sur l'appartenance des objets aux classes ; on cherche à inférer l'appartenance des objets aux classes à partir d'hypothèses sur la forme des classes et de la matrice de dissimilarités$\rightarrow$ apprentissage non supervisé

Cas 1 : apprentissage supervisé

- Les données :
  - Matrice de dissimilarités $D=(d_{ij})$, $d_{ij} \in \Delta$
  - n fonctions de masses (« étiquettes crédibilistes ») $m^{\Omega}\{c_i\}$, $i=1,\ldots,n$ sur $\Omega=\{\omega_1,\ldots,\omega_K\}$, relatives à la classe c_i de chaque objet o_i.
- Problème :
  - Déterminer une fonction de masse $m^{\Omega}\{c_0\}$ relative à la classe d'un nouvel objet dont on connaît les dissimilarités $d_{01},\ldots,d_{0n}$ par rapport aux n objets de la base.

Cas particuliers

- Etiquetage
  - Précis et certain :

    $m^{\Omega}\{c_i\}$ certaines : $\forall$ i, $\exists$ k, $m^{\Omega}\{c_i\}(\{\omega_k\})=1$
  - Ensembliste :

    $m^{\Omega}\{c_i\}$ catégoriques : $\forall$ i, $\exists$ $A \subseteq \Omega$, $m^{\Omega}\{c_i\}(A)=1$
  - Probabiliste :

    $m^{\Omega}\{c_i\}$ bayésiennes
  - Possibiliste :

    $m^{\Omega}\{c_i\}$ consonantes
  - ...

Cas 2 : apprentissage non supervisé

- Les données :
  - Matrice de dissimilarités $D=(d_{ij})$, $d_{ij} \in \Delta$
- Hypothèse :
  - chaque objet appartient à une classe parmi $\Omega=\{\omega_1,\ldots,\omega_K\}$
- Problème :
  - Trouver n masses $m^\Omega\{c_i\}$, $i=1,\ldots,n$ sur Ω, caractérisant l'appartenance des n objets aux K classes $\rightarrow$ partition crédale

Partition crédale : exemple

A	$m^{\Omega}\{c_1\}(A)$	$m^{\Omega}\{c_2\}(A)$	$m^{\Omega}\{c_3\}(A)$	$m^{\Omega}\{c_4\}(A)$	$m^{\Omega}\{c_5\}(A)$
$\emptyset$	0	0	0	0	0
$\{\omega_1\}$	0	0	0	0.2	0
$\{\omega_2\}$	0	1	0	0.4	0
$\{\omega_1, \omega_2\}$	0.7	0	0	0	0
$\{\omega_3\}$	0	0	0.2	0.4	0
$\{\omega_1, \omega_3\}$	0	0	0.5	0	0
$\{\omega_2, \omega_3\}$	0	0	0	0	0
Ω	0.3	0	0.3	0	1

Cas particuliers

- Chaque $m^{\Omega}\{c_i\}$ certaine $\rightarrow$ partition nette
- Chaque $m^{\Omega}\{c_i\}$ bayésienne $\rightarrow$ partition floue

$$u_{ik} = m^{\Omega}\{c_i\}(\{\omega_k\}), \quad \forall i, k$$

$$\sum_{k=1}^{K} u_{ik} = 1$$

- Chaque $m^{\Omega}\{c_i\}$ consonante (éléments focaux emboîtés) $\rightarrow$ partition possibiliste

$$u_{ik} = pl^{\Omega_i}(\{\omega_k\})$$

Résumé

	Connu	Inconnu
Problème 1	$D = (d_{ij})$ $m^{\Omega}\{c_1\}, \ldots, m^{\Omega}\{c_n\}$ $d_{01}, \ldots, d_{0n}$	$m^{\Omega}\{c_0\}$
Problème 2	$D = (d_{ij})$	$m^{\Omega}\{c_1\}, \ldots, m^{\Omega}\{c_n\}$

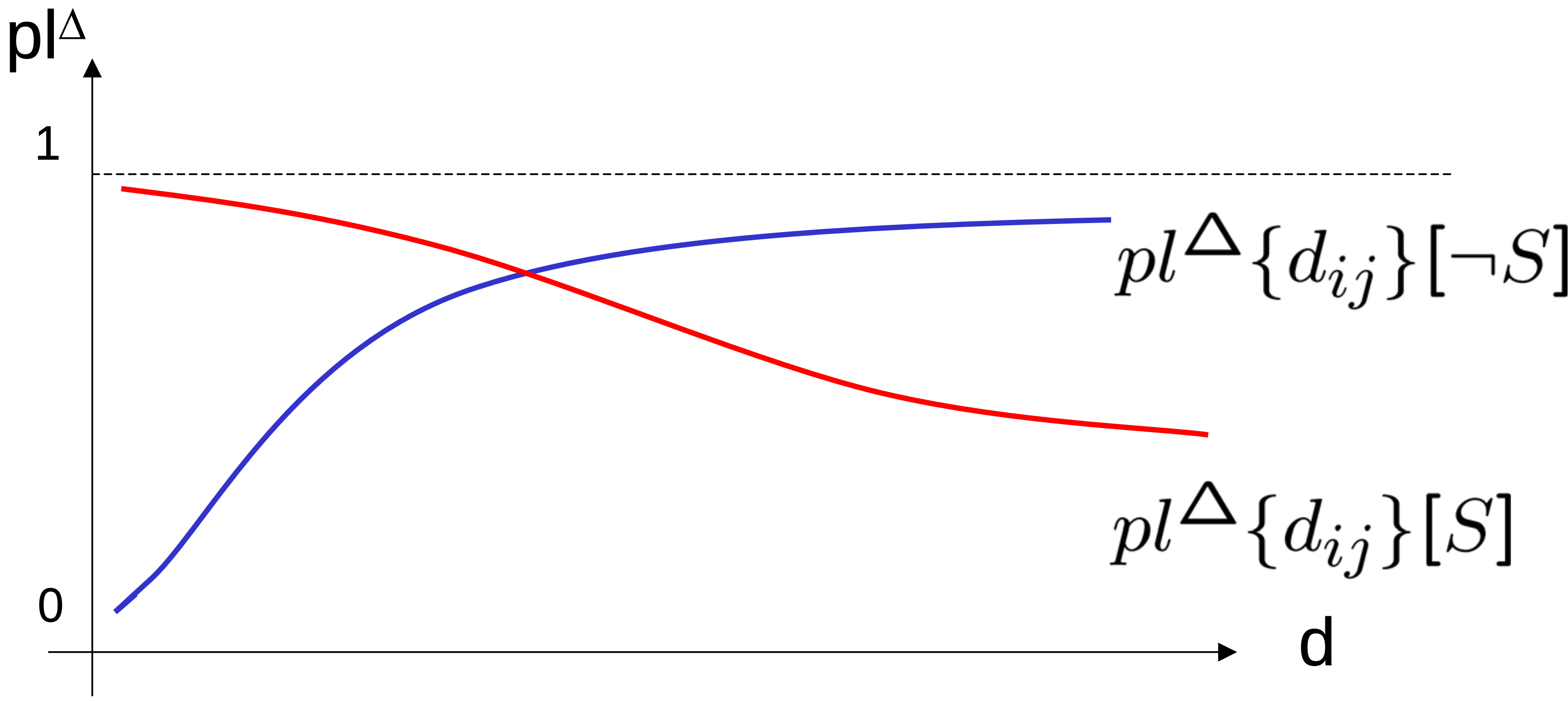
Solution du problème 1

- Nécessité de définir un modèle traduisant une connaissance a priori (même rudimentaire) sur les classes
- Soient deux objets o_i et o_j quelconques, de classes c_i et c_j. Soit S la proposition $c_i=c_j$
- On suppose connues les fonctions suivantes (identiques quels que soient i et j) :

$$d \longrightarrow pl^{\Delta}\{d_{ij}\}[S](d)$$ [fonction de plausibilité intra-classe]

$$d \longrightarrow pl^{\Delta}\{d_{ij}\}[\neg S](d)$$ [fonction de plausibilité inter-classe]

Fonctions de plausibilité inter et intra-classes



Application du GBT

- Soit $\Theta=\{S,\neg S\}$
- On connaît $pl^{\Delta}\{d_{ij}\}[S](d)$, $pl^{\Delta}\{d_{ij}\}[\neg S](d)$, $\forall d \in \Delta$
- On cherche $m^{\Theta}\{c_i,c_j\}[d_{ij}=d]$
- Théorème de Bayes généralisé (TBG) :

$$m^{\Theta}\{c_i,c_j\}[d_{ij}=d](A) = \prod_{\theta\in A} pl^{\Delta}\{d_{ij}\}[\theta](d)\times \prod_{\theta\in\overline{A}} (1-pl^{\Delta}\{d_{ij}\}[\theta](d)), \quad \forall A \subseteq \Theta$$

Application du GBT (suite)

$$
\begin{aligned}
m^{\Theta}\{c_i,c_j\}[d_{ij}=d](\{S\}) &= pl[S](d)(1-pl[\neg S](d))\\
m^{\Theta}\{c_i,c_j\}[d_{ij}=d](\{\neg S\}) &= pl[\neg S](d)(1-pl[S](d))\\
m^{\Theta}\{c_i,c_j\}[d_{ij}=d](\Theta) &= pl[\neg S](d)pl[S](d)\\
m^{\Theta}\{c_i,c_j\}[d_{ij}=d](\emptyset) &= (1-pl[S](d))(1-pl[\neg S](d))
\end{aligned}
$$

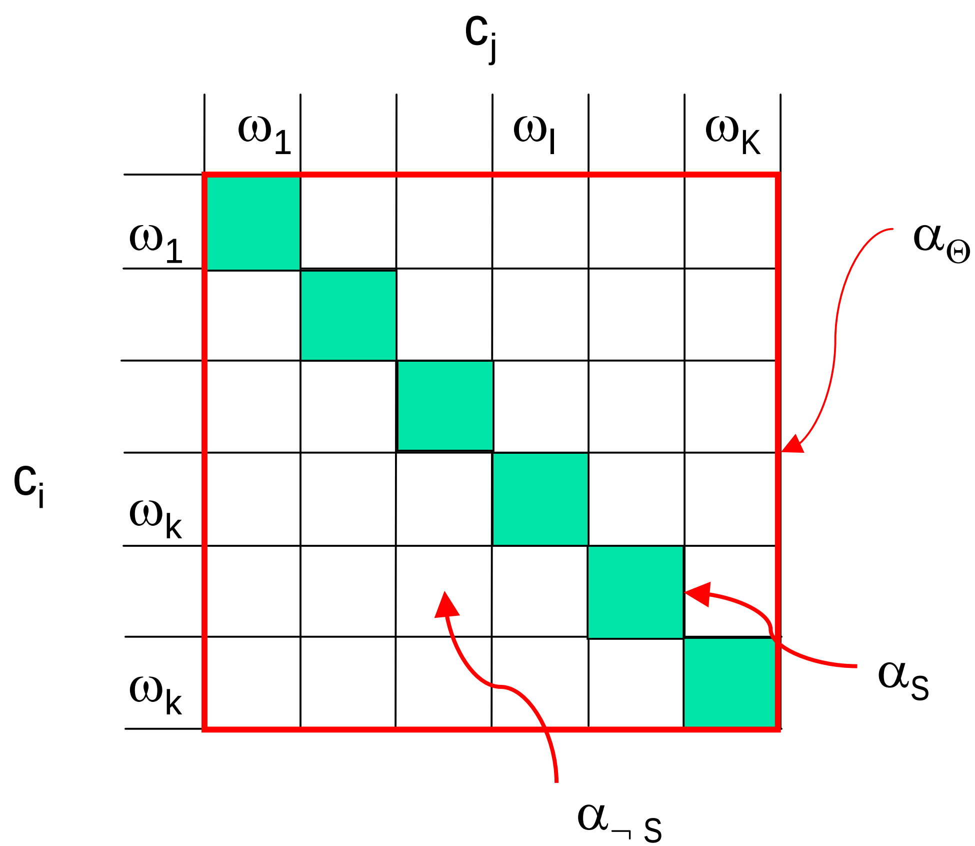
- Interprétation :
  - $pl[S](d) \approx 1$ et $pl[\neg S](d) \approx 0 \rightarrow m[d_{ij}=d]\{c_i,c_j\}(\{S\}) \approx 1$
  - $pl[S](d) \approx 0$ et $pl[\neg S](d) \approx 1 \rightarrow m[d_{ij}=d]\{c_i,c_j\}(\{\neg S\}) \approx 1$
  - $pl[S](d) \approx 1$ et $pl[\neg S](d) \approx 1 \rightarrow m[d_{ij}=d]\{c_i,c_j\}(\Theta) \approx 1$
  - $pl[S](d) \approx 0$ et $pl[\neg S](d) \approx 0 \rightarrow m[d_{ij}=d]\{c_i,c_j\}(\emptyset) \approx 1$

Extension à Ω^2

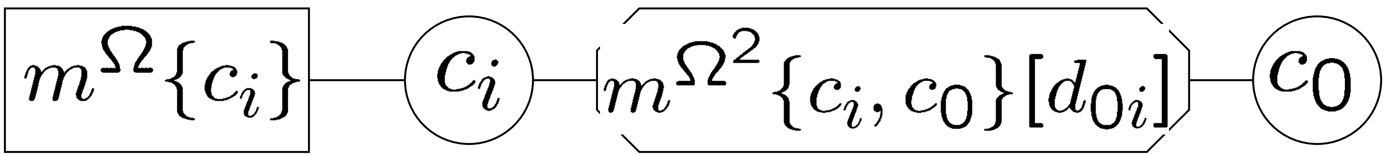
- $\Theta=\{S, \neg S\}$ est un grossissement de Ω^2
  $S \equiv \{(\omega_1,\omega_1),\ldots,(\omega_K,\omega_K)\} \subseteq \Omega^2$
- Extension de $m^\Theta\{c_i,c_j\}[d]$ à Ω^2

$$
\begin{aligned}
m^{\Theta}\{c_i,c_j\}[d](\{S\}) &\longrightarrow m^{\Omega^2}\{c_i,c_j\}[d](S) &\triangleq \alpha_S\\
m^{\Theta}\{c_i,c_j\}[d](\{\neg S\}) &\longrightarrow m^{\Omega^2}\{c_i,c_j\}[d](\neg S) &\triangleq \alpha_{\neg S}\\
m^{\Theta}\{c_i,c_j\}[d](\Theta) &\longrightarrow m^{\Omega^2}\{c_i,c_j\}[d](\Omega^2) &\triangleq \alpha_{\Theta}\\
m^{\Theta}\{c_i,c_j\}[d](\emptyset) &\longrightarrow m^{\Omega^2}\{c_i,c_j\}[d](\emptyset) &\triangleq \alpha_{\emptyset}
\end{aligned}
$$

Extension à Ω^2 (suite)



Combinaison avec $m^{\Omega}\{c_i\}$



$$m^{\Omega}\{c_0\}[e_i] = \left(m^{\Omega^2}\{c_0, c_i\}[d_{0i}] \textcircled{$\cap$} m^{\Omega}\{c_i\} \right)^{\downarrow \Omega}\{c_0\}$$

avec $e_i = (d_{0i}, m^{\Omega}\{c_i\})$.

Résultat

$$\boxed{m^{\Omega^2}\{c_0\}[e_i] = f(m^{\Omega^2}\{c_i\}, \alpha_S, \alpha_{\neg S}, \alpha_\ominus, \alpha_\emptyset)}$$

$$m^\Omega\{c_0\}[e_i](\emptyset) = \alpha_\emptyset(1 - m^\Omega\{c_i\}(\emptyset)) + m^\Omega\{c_i\}(\emptyset)$$

$$m^\Omega\{c_0\}[e_i](A) = \alpha_S m^\Omega\{c_i\}(A), \quad \forall A, 0 < |A| < |\Omega| - 1$$

$$m^\Omega\{c_0\}[e_i](A) = \alpha_S m^\Omega\{c_i\}(A) + \alpha_{\neg S} m^\Omega\{c_i\}(\overline{A}), \quad \forall A, |A| = |\Omega| - 1$$

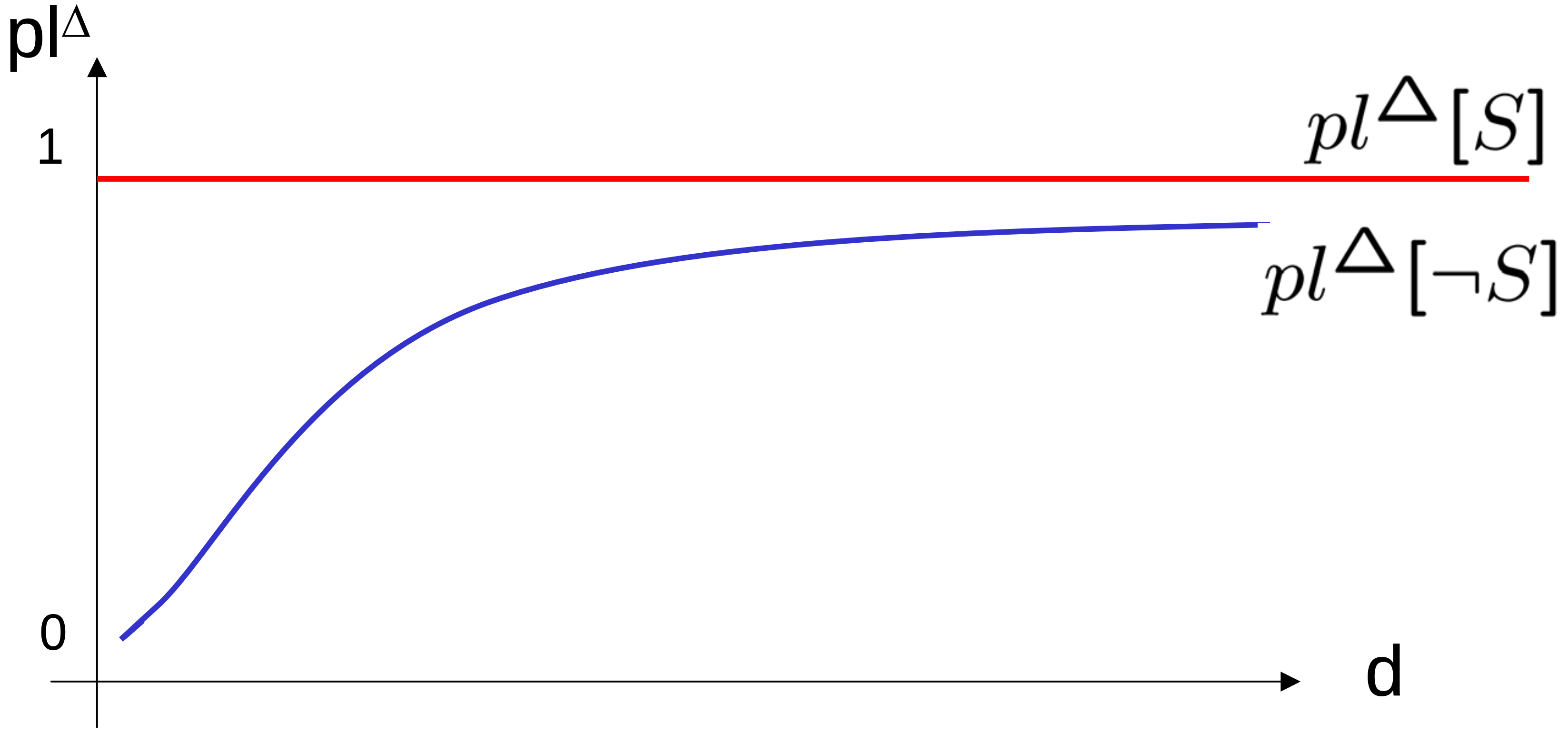
$$m^\Omega\{c_0\}[e_i](\Omega) = \alpha_\ominus(1 - m^\Omega\{c_i\}(\emptyset)) + \alpha_S m^\Omega\{c_i\}(\Omega) + \alpha_{\neg S} \sum_{B \subseteq \Omega, |B| > 1} m^\Omega\{c_i\}(B)$$

Combinaison sur l'ensemble d'apprentissage

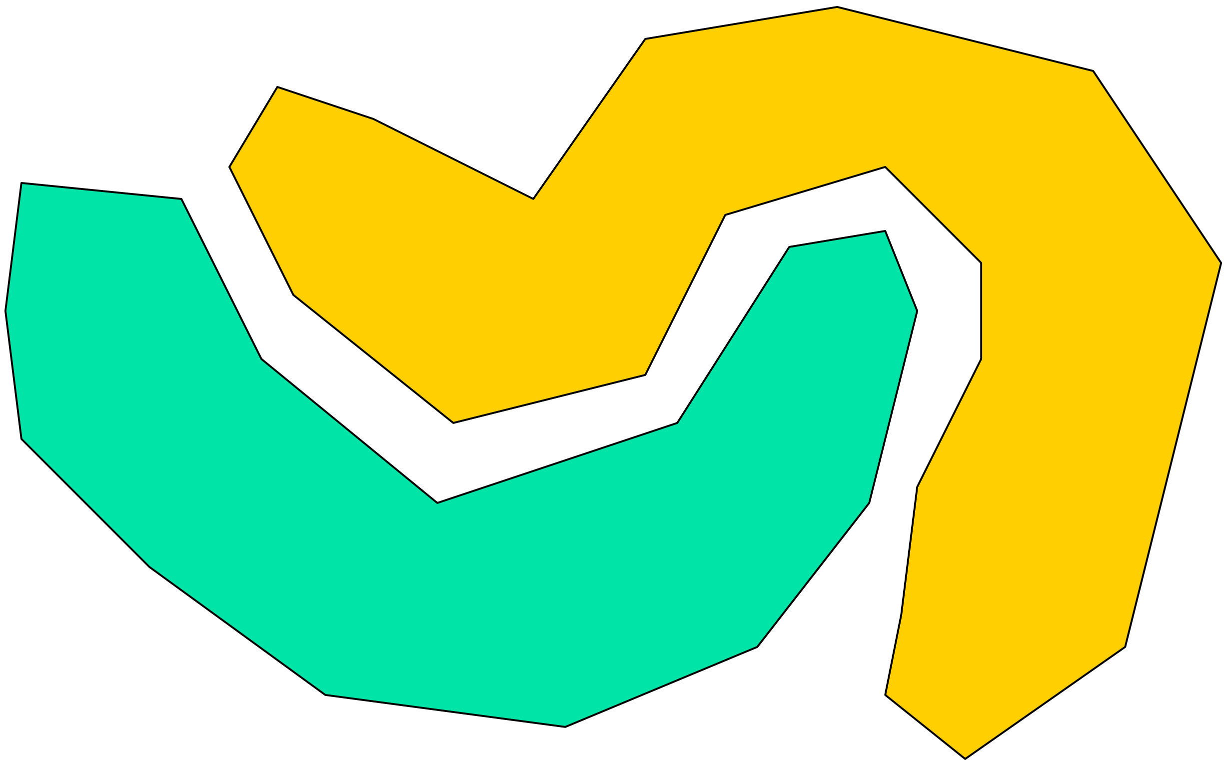
$$m^{\Omega}\{c_0\}[\mathcal{A}] = \bigcirc\!\!\!\!\cap_{i=1}^{n} m^{\Omega}\{c_0\}[e_i]$$

avec $\mathcal{A} = (e_1, \ldots, e_n)$.

Cas particulier 1



Cas particulier 1 (suite)



Cas particulier 1 (suite)

$$m^{\Omega}\{c_0\}[e_i](A) = \alpha_S m^{\Omega}\{c_i\}(A), \quad \forall A, 0 < |A| < |\Omega|$$

$$m^{\Omega}\{c_0\}[e_i](\Omega) = 1 - \alpha_S + \alpha_S m^{\Omega}\{c_i\}(\Omega)$$

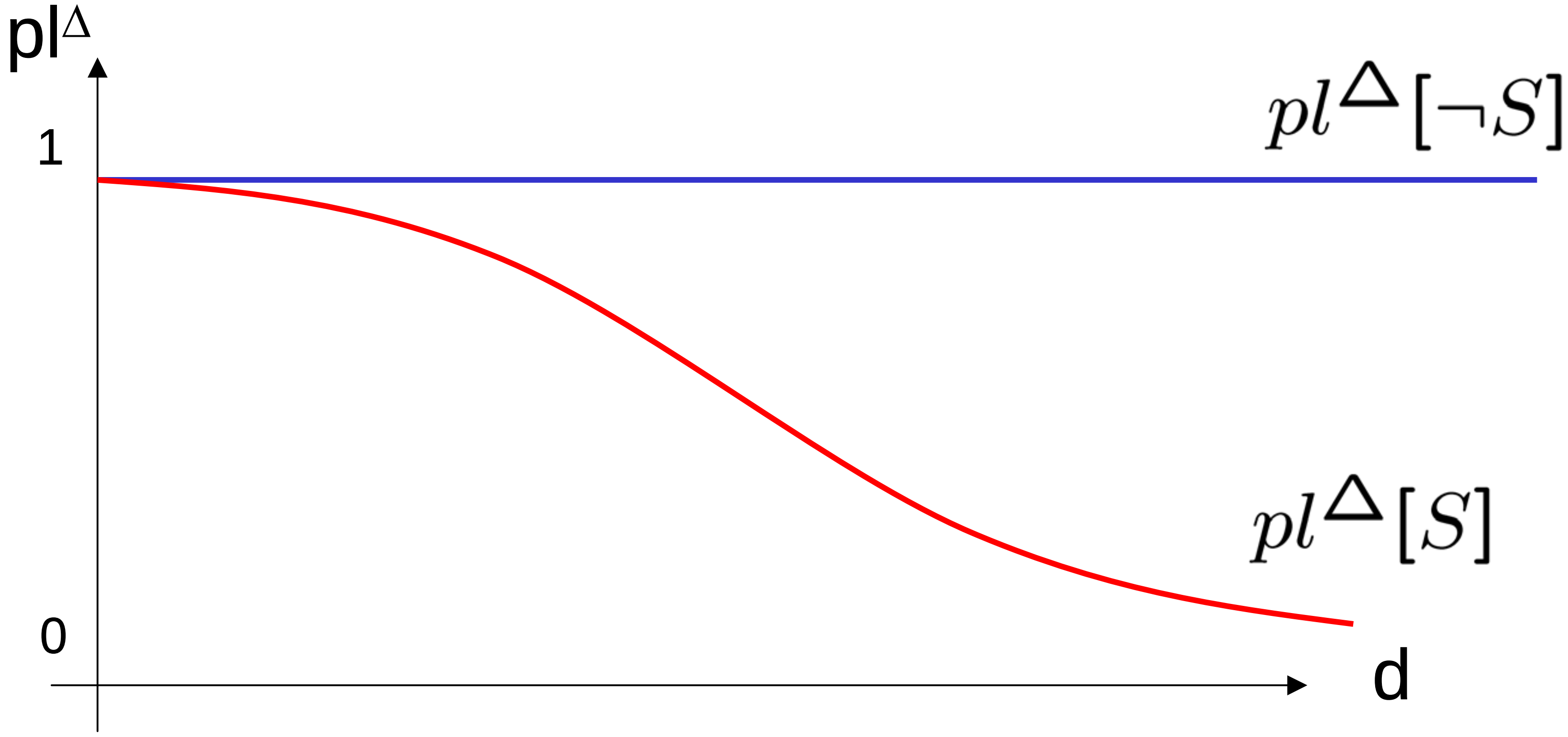
avec $\alpha_S = 1 - pl^{\Delta}[\neg S](d_{0i})$.

Coefficient d'affaiblissement = $pl^{\Delta}[\neg S\}](d_{0i})$ fonction croissante de d_{0i}
$\rightarrow$ les exemples les plus proches (similaires) sont les plus informatifs

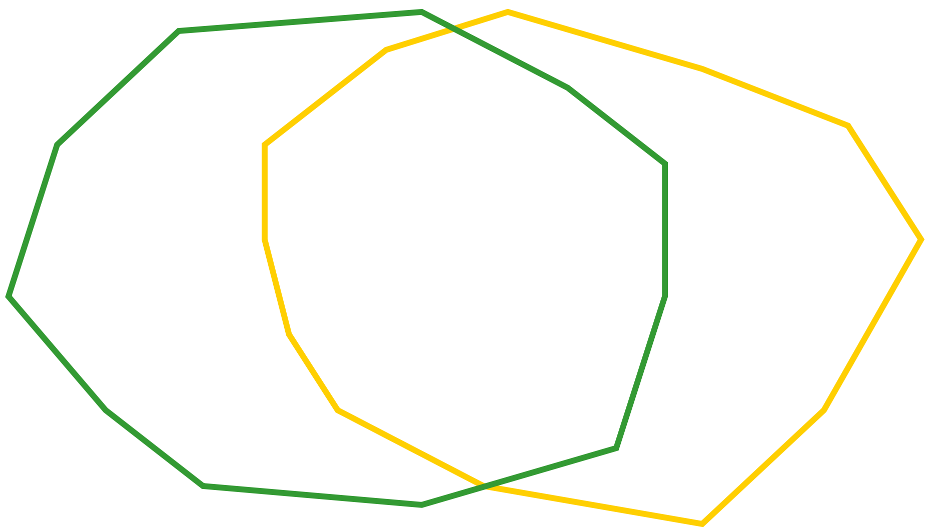
Méthode introduite "empiriquement" dans :

T. Denoeux. A *k*-nearest neighbor classification rule based on Dempster-Shafer theory. *IEEE Transactions on Systems, Man and Cybernetics*, 25(05):804-813, 1995.

Cas particulier 2



Cas particulier 2 (suite)



Cas particulier 2 (suite)

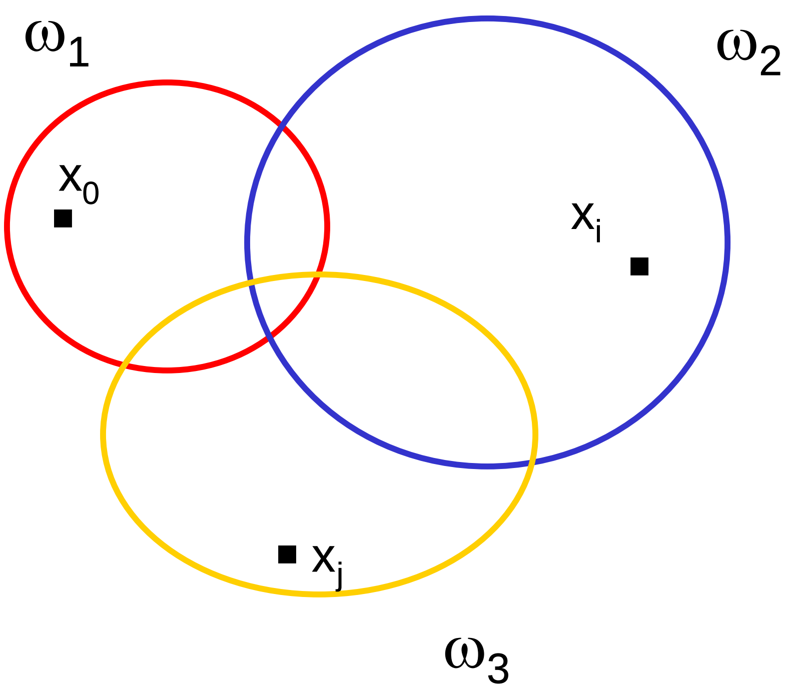
$$m^{\Omega}\{c_0\}[e_i](\overline{\{\omega_k\}}) = \alpha_{\neg S} m^{\Omega}\{c_i\}(\{\omega_k\}), \quad \forall \omega_k \in \Omega$$

$$m^{\Omega}\{c_0\}[e_i](\Omega) = 1 - \alpha_{\neg S} + \alpha_{\neg S} \sum_{B \subseteq \Omega, |B|>1} m^{\Omega}\{c_i\}(B)$$

avec $\alpha_{\neg S} = 1 - pl^{\Delta}[S](d_{0i})$.

Coefficient « d'affaiblissement » = $pl^{\Delta}[S\}](d_{0i})$ fonction décroissante de d_{0i} $\rightarrow$ les exemples les plus éloignés sont les plus informatifs

Exemple



Mise en œuvre pratique

- Définir pour $pl^{\Delta}[S]$ et $pl^{\Delta}[\neg S]$ des familles de fonctions paramétrées
- Exemple :
  - $pl^{\Delta}[S](d)=\exp(-\gamma_S\, d)$
  - $pl^{\Delta}[\neg S](d)=1-\exp(-\gamma_{\neg S}\, d)$
- Optimisation des paramètres par minimisation d'une fonction d'erreur

$$\boxed{E(\gamma_S,\gamma_{\neg S}) = \sum_{i=1}^{n} C(m^{\Omega}\{c_i\}[\mathcal{A}\setminus e_i], m^{\Omega}\{c_i\})}$$

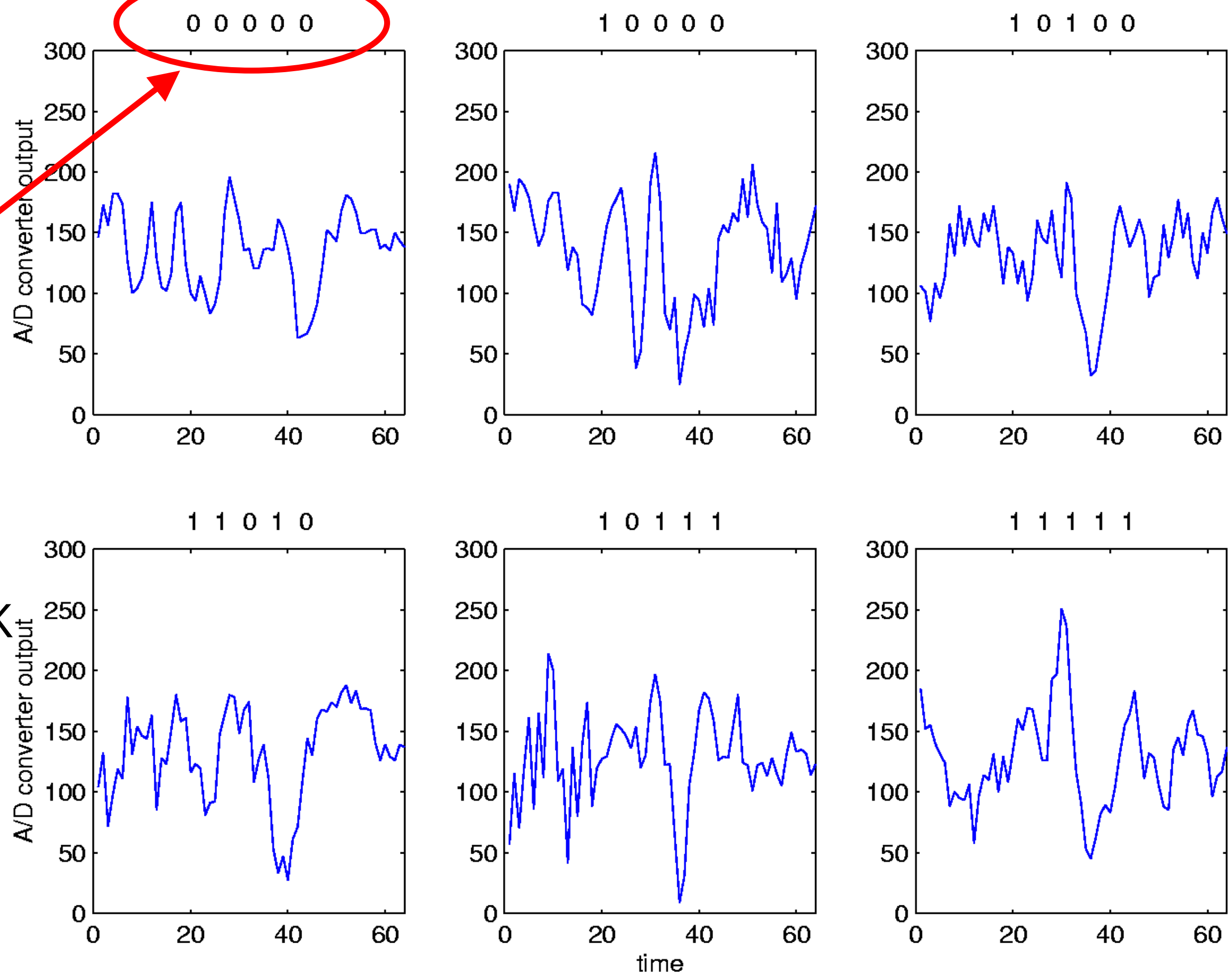
Exemple : classification de signaux EEG

- Pb : discriminer les complexes K du signal de fond dans des signaux EEG enregistrés pendant le sommeil (Richard, 1998).
- Complexe K = forme transitoire, utile pour l'étiquetage des stades de sommeil et le diagnostic en psychiatrie.
- Problèmes :
  - Absence de critère objectif : étiquetage des données par un panel d'experts ;
  - Absence de consensus entre les experts.

Détection de Complexes K

étiquetage subjectif par 5 experts

0= onde delta
1 = complexe K



Données EEG

- K=2 classes, 64 variables ($\rightarrow$ distance euclidienne)
- données étiquetées par 5 experts
- n=200 exemples d'apprentissage, 300 exemples de test.

k	k-ppv	k-ppv pondérés	MCT (étiquetage majoritaire)	MCT (étiquetage incertain)
9	0.30	0.30	0.31	0.27
11	0.29	0.30	0.29	0.26
13	0.31	0.30	0.31	0.26

Solution du problème 2 (apprentissage non supervisé)

- Données : matrice $D=(d_{ij})$ de dissimilarités entre n objets (données de proximités)
- Objectif : Associer à chaque objet une fonction de masse $m^{\Omega}\{c_i\}$ sur un ensemble de classes $\Omega=\{\omega_1,\ldots,\omega_K\}$ défini a priori.
- Les n fonctions de masse $\{m^{\Omega}\{c_i\}, i=1,..,n\}$ constituent une partition crédale des n objets.

Principe de construction

- Soit M={$m^\Omega\{c_i\}$, i=1,..,n} une partition crédale de n objets.
- $\forall$ i,j, la fonction de masse jointe sur l'espace produit Ω^2 est obtenue par :

$$m^{\Omega^2}\{c_i,c_j\} = m^{\Omega}\{c_i\} \circledcap m^{\Omega}\{c_i\}$$

- On a :

$$\begin{aligned}
pl^{\Omega^2}\{c_i,c_j\}(S) &= \sum_{\{A\times B\subseteq \Omega^2 \mid (A\times B)\cap S\neq\emptyset\}} m^{\Omega^2}\{c_i,c_j\}(A\times B) \\
&= \sum_{A\cap B\neq\emptyset} m^{\Omega}\{c_i\}(A)\cdot m^{\Omega}\{c_j\}(B) \\
&= 1 - K_{ij}
\end{aligned}$$

degré de conflit entre $m^\Omega\{c_i\}$ et $m^\Omega\{c_j\}$

Principe de construction (suite)

- Définition : Soit $M=\{m^\Omega\{c_i\}, i=1,..,n\}$ une partition crédale de n objets. M est compatible avec la matrice $D=(d_{ij})$ des dissimilarités ssi, $\forall$ i,j,i',j' :

$$d_{ij} > d_{i'j'} \Rightarrow K_{ij} \geq K_{i'j'}$$

« Plus deux objets sont proches, plus il est plausible qu'ils appartiennent à la même classe »

- Recherche d'une partition crédale approximativement compatible par minimisation de :

$$I(M,a,b) \triangleq \sum_{i<j} \frac{(aK_{ij} + b - d_{ij})^2}{d_{ij}}$$

- Analogie avec le positionnement multidimensionnel (*multidimensional scaling*)

Mise en oeuvre

- Nombre de paramètres à optimiser : n 2^K, pour n(n-1)/2 dissimilarités.
- Solutions :
  - réduction des éléments focaux aux singletons $\{\omega_k\}$ (k=1,...,K), $\emptyset$, Ω
  - Recherche d'une partition crédale la plus « informative » (précise et certaine) possible par ajout d'un terme d'entropie:

$$I' = I + \lambda \sum_{i=1}^{n} H(m^{\Omega}\{c_i\})$$

Choix d'un critère d'entropie

- Critère de Pal et Bezdek

$$H(m) = \sum_{\{A \in 2^\Omega \setminus \{\emptyset\}, m(A)>0\}} m(A) \log_2 \left(\frac{|A|}{m(A)} \right) + m(\emptyset) \log_2 \left(\frac{|\Omega|}{m(\emptyset)} \right)$$

- H favorise l'allocation de la masse à un petit nombre d'éléments focaux de faible cardinalité.

Exploitation des résultats

partition crédale
$\{m^\Omega\{c_i\}, i=1,\ldots,n\}$

$m^\Omega\{c_i\}(\emptyset) \approx 1 \rightarrow$ « outlier »
$m^\Omega\{c_i\}(\Omega) \approx 1 \rightarrow$ « inlier »

Transformation pignistique

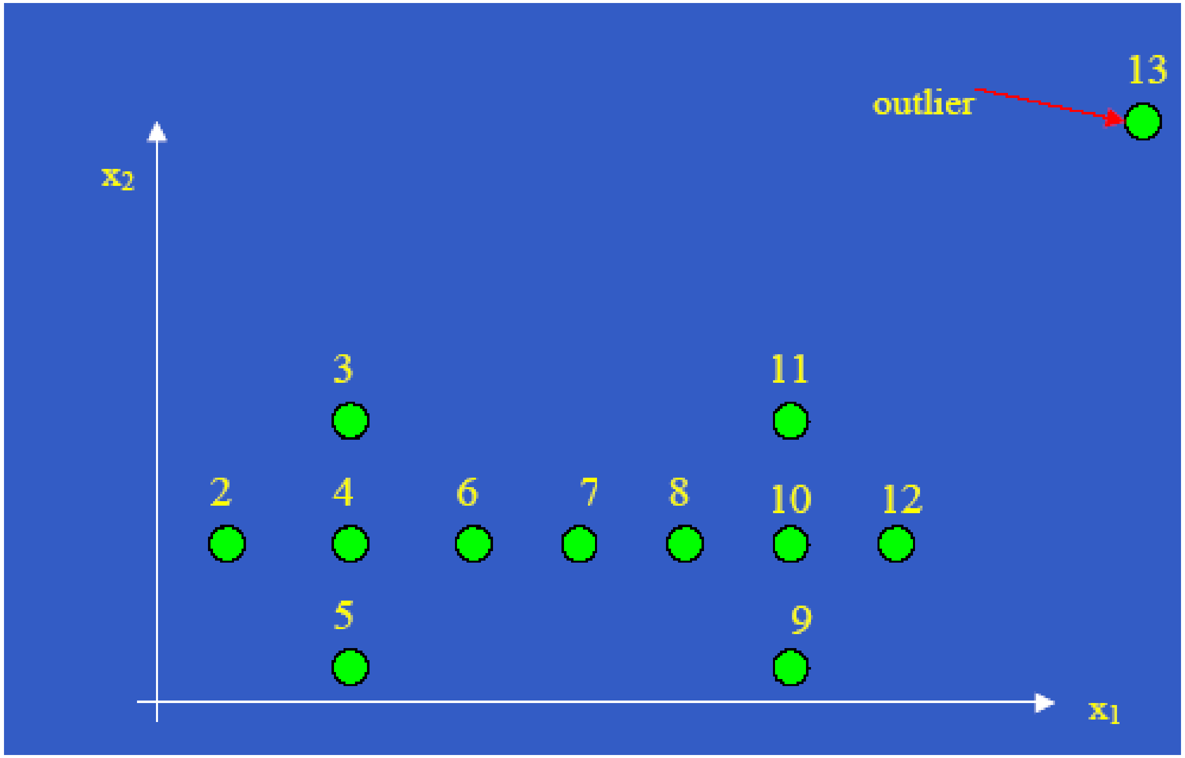
$$BetP(\omega) = \sum_{\omega \in A} \frac{m(A)}{(1 - m(\emptyset))|A|}$$

partition floue
$\{BetP^\Omega\{c_i\}, i=1,\ldots,n\}$

Maximum de probabilité pignistique

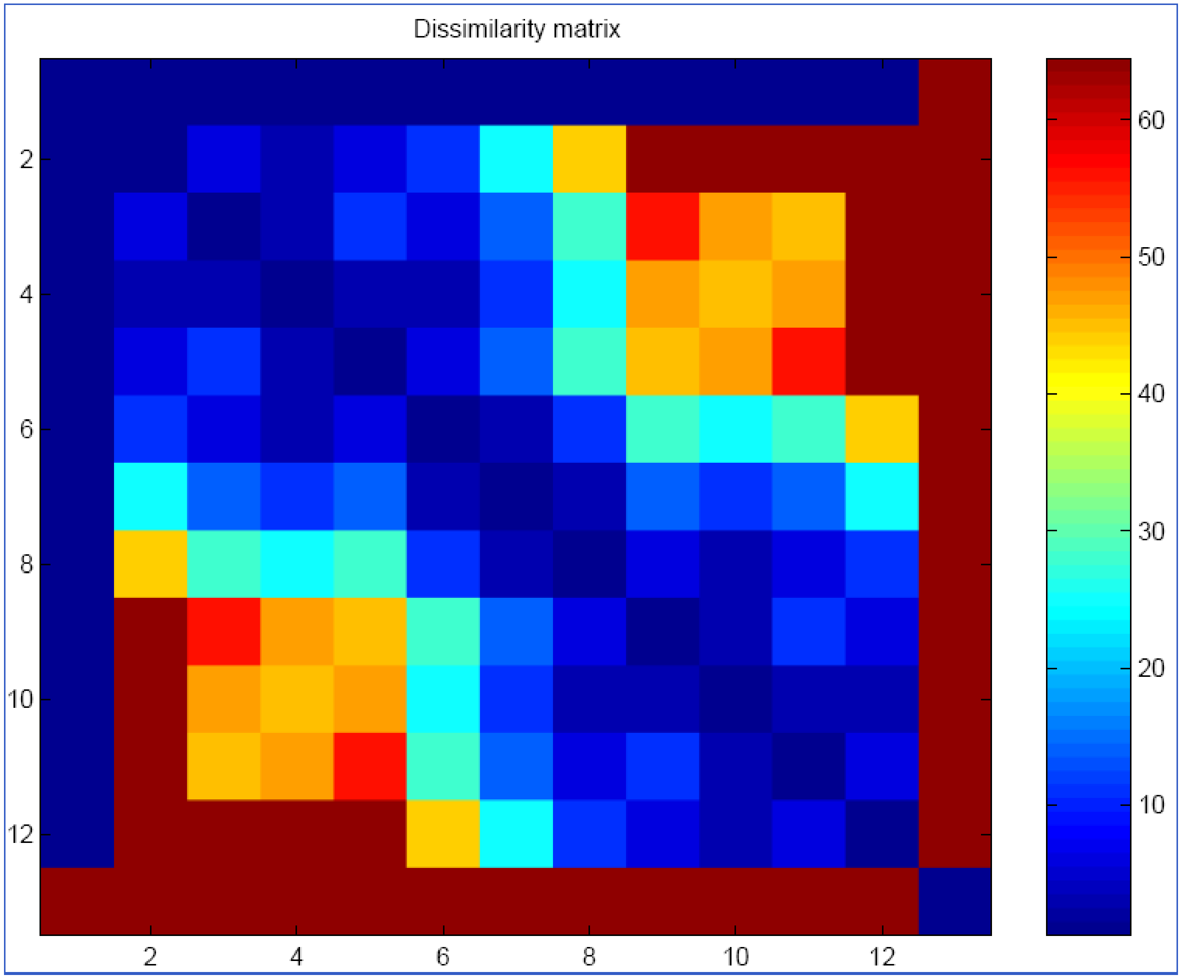
partition nette
$u_{ik} \in \{0,1\}$
$i=1,\ldots,n,\ k=1,\ldots,K$

Exemple 1

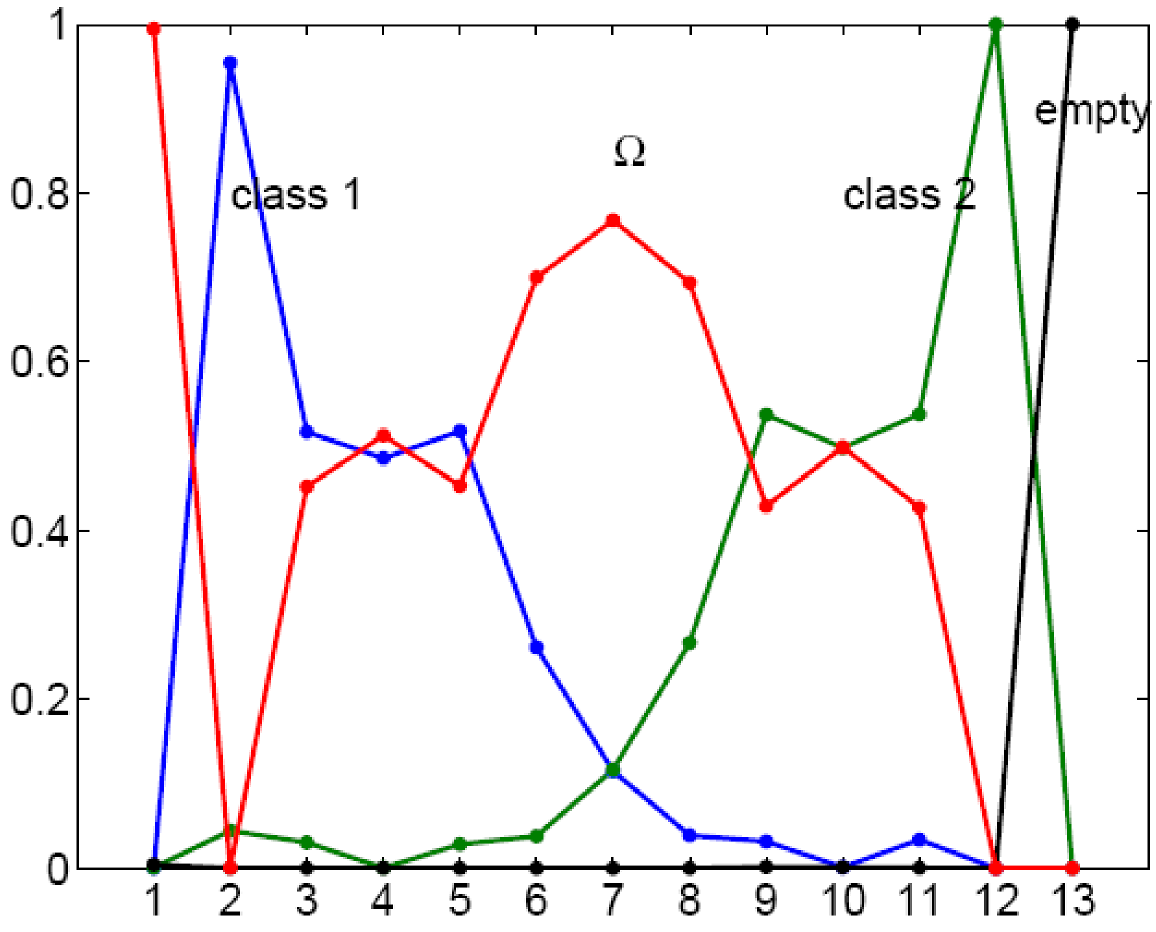


+ objet 1 similaire à tous les autres objets (« inlier »)

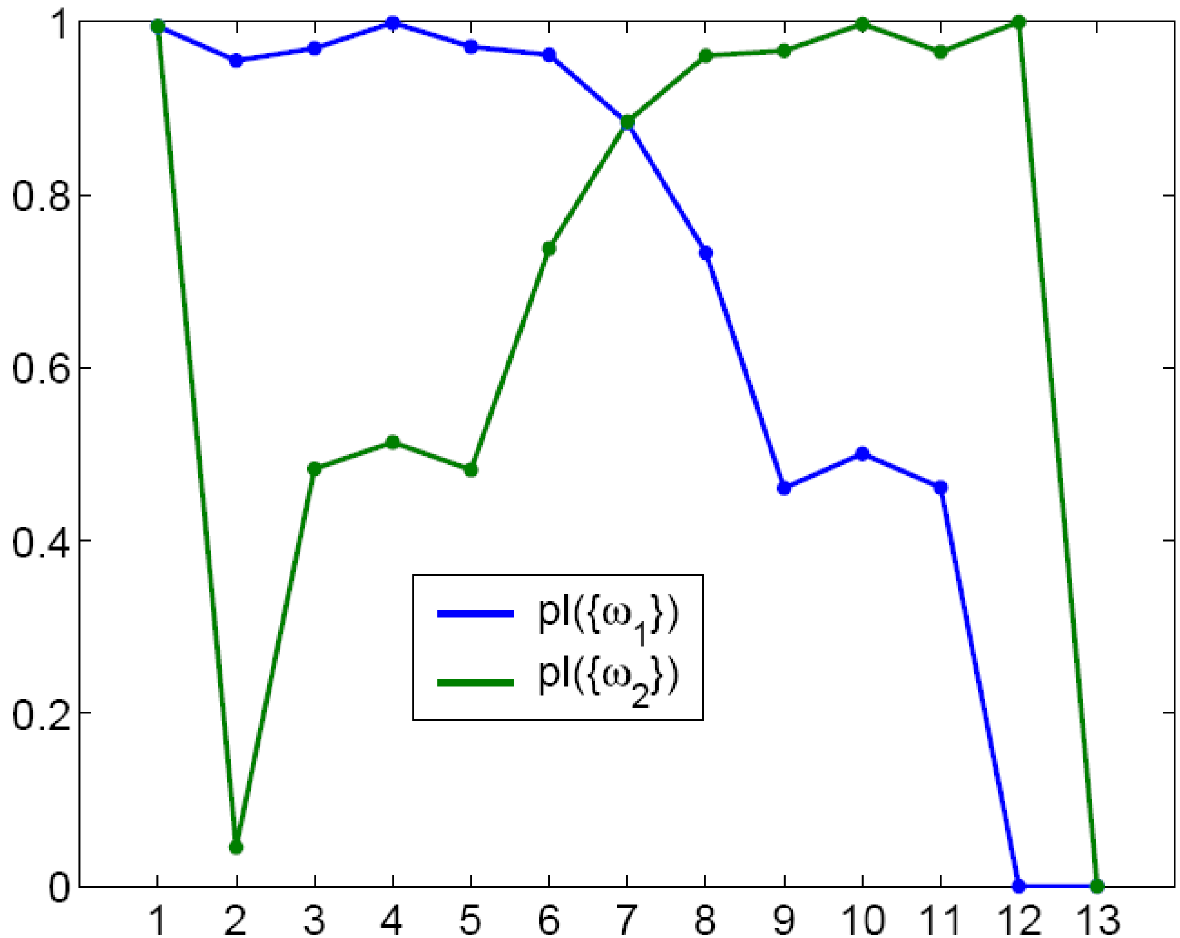
Exemple 1: matrice de dissimilarités



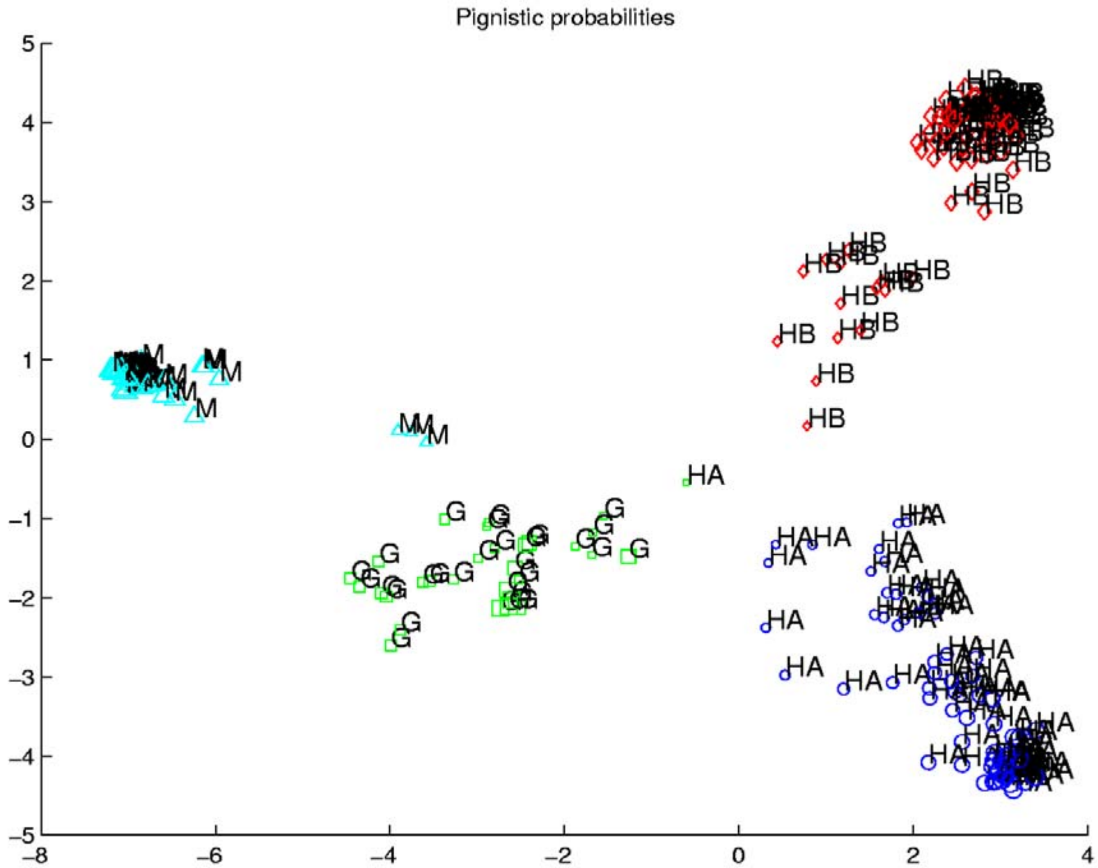
Exemple 1 : résultat (masses)



Exemple 1 : résultat (plausibilités)



Exemple 2 : protéines

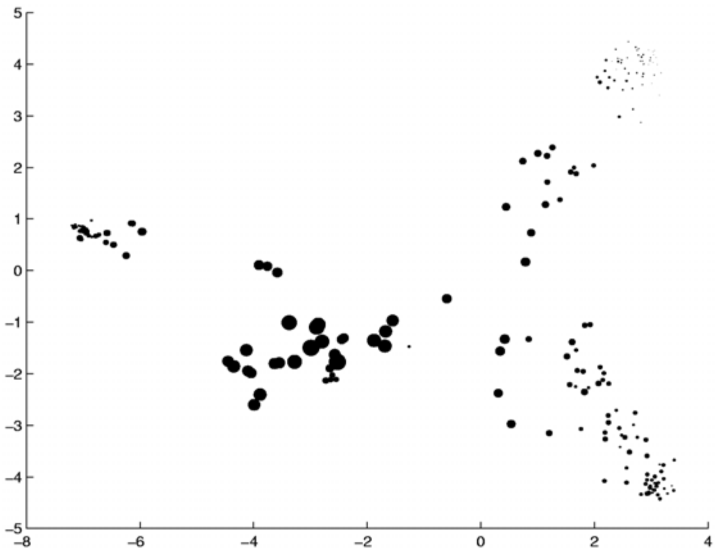


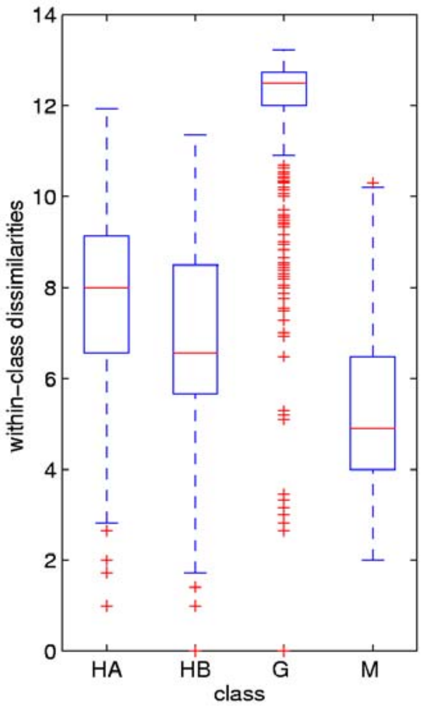
Données :

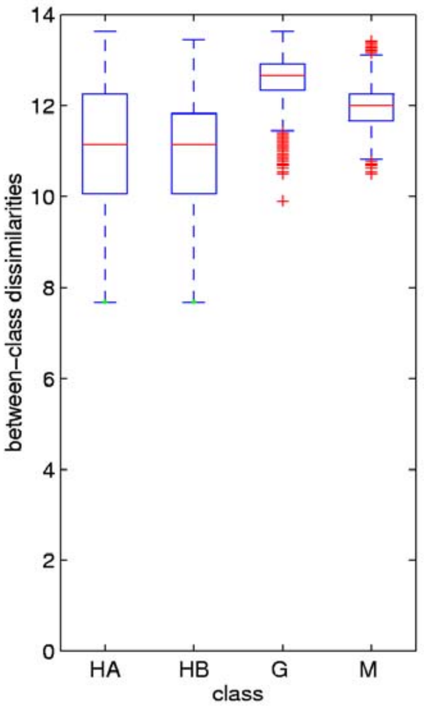
- comparaison structurelle de 213 protéines,
- 4 classes (hémogolbine-α et β, myoglobines, globines hétérogènes).

Protéines (suite)

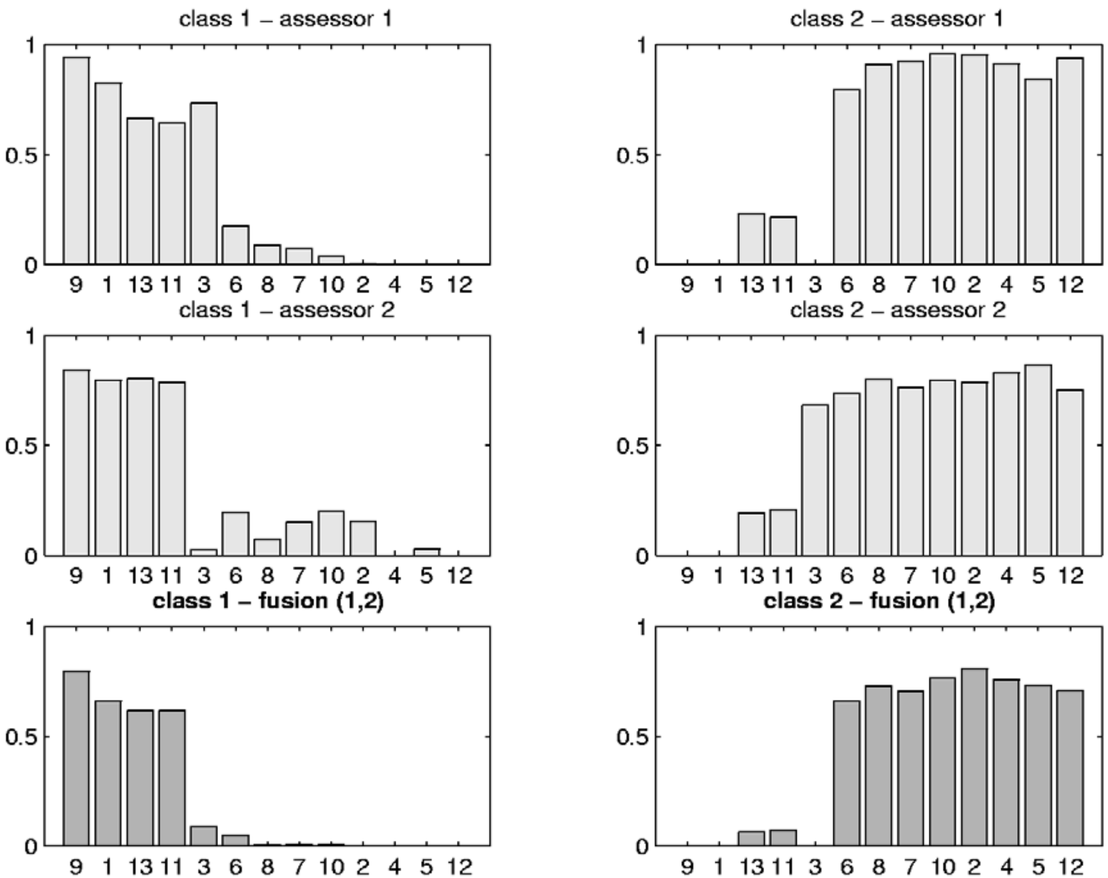
$m^{\Omega}\{c_i\}(\emptyset)$



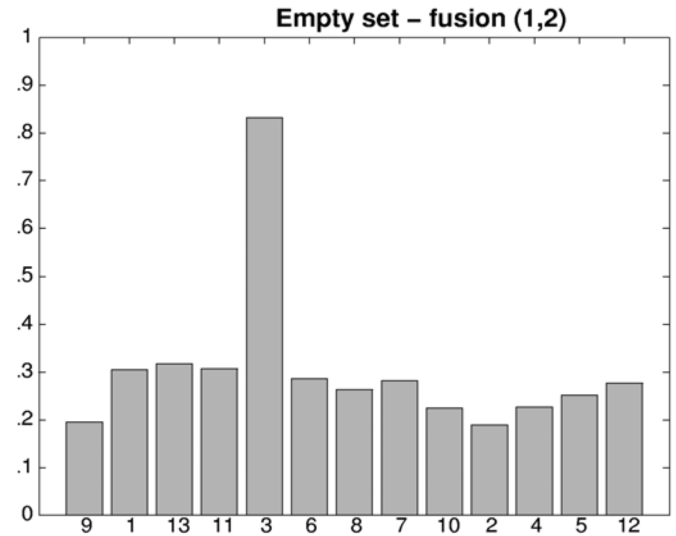




Exemple 3 : analyse sensorielle



13 produits,
4 descripteurs,
2 juges



Conclusion

- Construction de fonctions de croyance à partir de dissimilarités dans deux cas :
  - Apprentissage complètement ou partiellement supervisé
  - Apprentissage non supervisé
- Hypothèses explicite sur la « forme des classes » :
  - Fonctions pl[S] et pl[$\neg$ S] en apprentissage supervisé
  - Hypothèse de monotonie entre d_{ij} et pl($\neg$ S) en apprentissage non supervisé.
- Le formalisme des fonctions de croyance permet :
  - D'exploiter en apprentissage une information incertaine et imprécise (étiquetage imprécis par exemple)
  - De traduire le résultat d'une analyse statistique des données sous forme incertaine et imprécise (partition crédale)