

Intégration de connaissance a priori en classification automatique

V. Antoine¹

B. Quost² M.-H. Masson² T. Denœux² N. Labroche³

¹Limos, UMR CNRS 6158, Clermont-Ferrand

²Heudiasyc, UMR CNRS 6599, Compiègne

³LI, EA 6300, Tours

Séminaire LIFO, février 2015

Plan

- Introduction
 - Classification automatique
 - Classification sous contraintes
 - Motivations
- Contributions
 - ECM-Mah
 - SECM
 - CECM
- Expérimentations
 - Comportement des algorithmes
 - Applications
- Conclusion

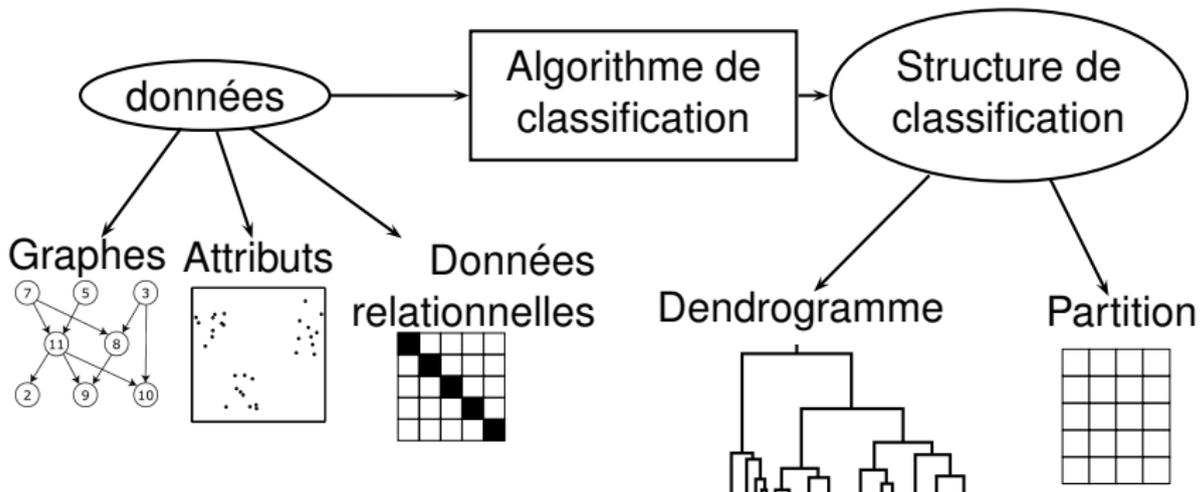
Plan

- Introduction
 - Classification automatique
 - Classification sous contraintes
 - Motivations
- Contributions
 - ECM-Mah
 - SECM
 - CECM
- Expérimentations
 - Comportement des algorithmes
 - Applications
- Conclusion

Introduction

Classification automatique

Grouper N individus en classes selon une notion de similarité



Types de partitions

- Partition dure (*k*-means)

- un objet x_i appartient à une classe de manière exclusive



- Partition floue (FCM)

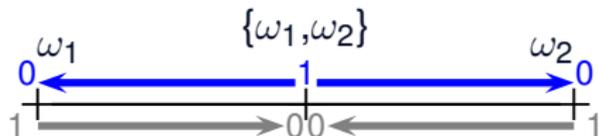
- degré d'appartenance d'un objet x_i à chaque classe ω_k : U_{ik}



- Partition crédale (ECM, EVCLUS)

- degré de croyance d'un objet x_i à chaque sous-ensemble

A_j de $\Omega = \{\omega_1, \dots, \omega_c\}$: m_{ij}



Types de partitions

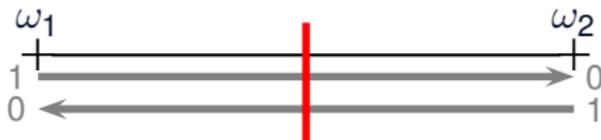
- Partition dure (k -means)

- un objet x_i appartient à une classe de manière exclusive



- Partition floue (FCM)

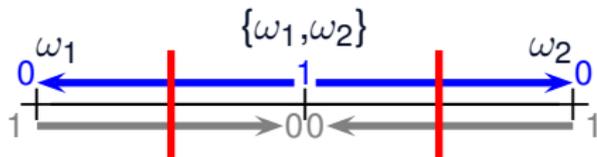
- degré d'appartenance d'un objet x_i à chaque classe $\omega_k : U_{ik}$



- Partition crédale (ECM, EVCLUS)

- degré de croyance d'un objet x_i à chaque sous-ensemble

A_j de $\Omega = \{\omega_1, \dots, \omega_c\} : m_{ij}$



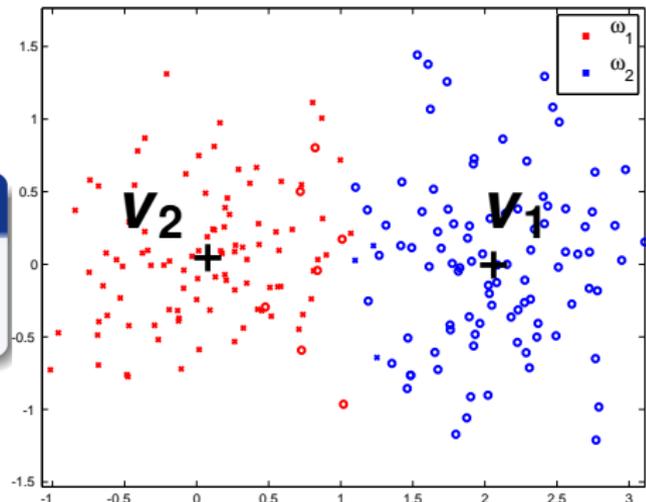
Algorithme des c-moyennes floues (FCM)

Partition floue

$$U = (u_{ik}), \quad \sum_{k=1}^C u_{ik} = 1 \text{ et } u_{ik} \geq 0 \quad \forall i \in \{1 \dots n\}, k \in \{1 \dots c\}$$

Modèle géométrique

Chaque classe ω_k est représentée par un centre v_k



Algorithme des c-moyennes floues (FCM)

Modèle géométrique

Distance

- Euclidienne : $d_{ik}^2 = \|\mathbf{x}_i - \mathbf{v}_k\|$
- Mahalanobis, méthode GK [1] : $d_{ik}^2 = (\mathbf{x}_i - \mathbf{v}_k)^\top \mathbf{S}_k (\mathbf{x}_i - \mathbf{v}_k)$

Fonction objectif

$$J_{FCM}(U, V, S) = \sum_{i=1}^N \sum_{k=1}^C u_{ik}^\beta d_{ik}^2$$

Optimisation alternée

$$\begin{array}{ccc} \min_{u_{ik}} J_{FCM} & \rightarrow & \min_{\mathbf{v}_k} J_{FCM} \\ & \swarrow & \searrow \\ & \min_{\mathbf{S}_k} J_{FCM} & \end{array}$$



[1] D. Gustafson & W. Kessel, *Fuzzy clustering with a fuzzy covariance matrix*, IEEE Conference on Decision and Control, 1978

Théorie des fonctions de croyance

Soit Y une variable prenant des valeurs dans un ensemble fini Ω .

Fonction de masse m

$$\sum_{A \subseteq \Omega} m(A) = 1$$

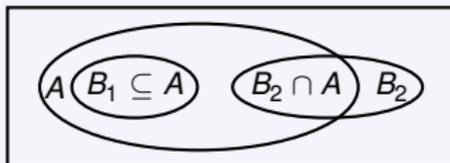
- $m(A)$: degré de croyance spécifique que $Y \in A$
- Si $m(A) > 0$ alors m est appelé élément focal

Notions dérivées

Fonction de plausibilité

Degré *Potentiel* de croyance qui peut être attribué à A :

$$pl(A) = \sum_{B \cap A \neq \emptyset} m(B), \quad \forall A \subseteq \Omega$$



Prise de décision : la transformation pignistique

Espace des fonctions de croyance \rightarrow espace probabiliste

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

Partition crédale

Problème de classification

- ⇒ Ω : ensemble des classes $\{\omega_1, \dots, \omega_c\}$
- ⇒ Y : la classe réelle de l'objet o_i
- ⇒ \mathbf{m}_i : connaissance partielle sur la classe de l'objet i
- ⇒ $\mathbf{M} = (\mathbf{m}_i)$: partition crédale

Exemple

A	m_1	m_2	m_3	m_4
\emptyset	0	0	0	1
$\{\omega_1\}$	1	0.3	0	0
$\{\omega_2\}$	0	0.7	0	0
$\{\omega_1, \omega_2\}$	0	0	1	0

Algorithmes évidentiels

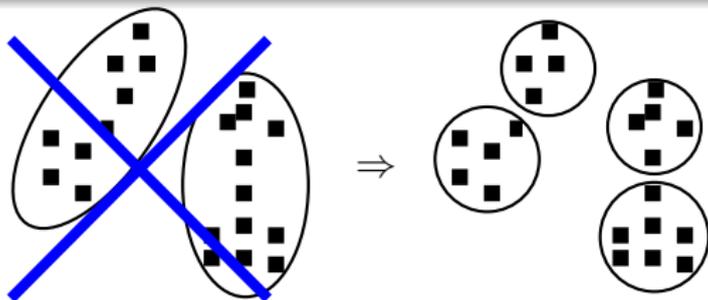
- modèle vectoriel : **ECM**
- modèle relationnel : **EVCLUS**

Classification sous contraintes

Type de contraintes

Intégrer des contraintes dans la classification automatique :

- au niveau du modèle
 - classes équilibrées
 - information négative : un modèle rejeté
- au niveau des classes
- au niveau des objets

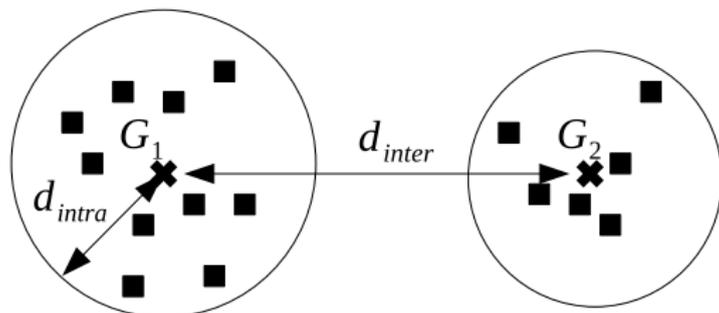


Classification sous contraintes

Type de contraintes

Intégrer des contraintes dans la classification automatique :

- au niveau du modèle
 - classes équilibrées
 - information négative : un modèle rejeté
- au niveau des classes
- au niveau des objets



Classification sous contraintes

Type de contraintes

Intégrer des contraintes dans la classification automatique :

- au niveau du modèle
 - classes équilibrées
 - information négative : un modèle rejeté
- au niveau des classes
- au niveau des objets

Must-Link



Cannot-Link



- \mathcal{M} ensemble de contraintes Must-Link
- \mathcal{C} ensemble de contraintes Cannot-Link

Classification sous contraintes

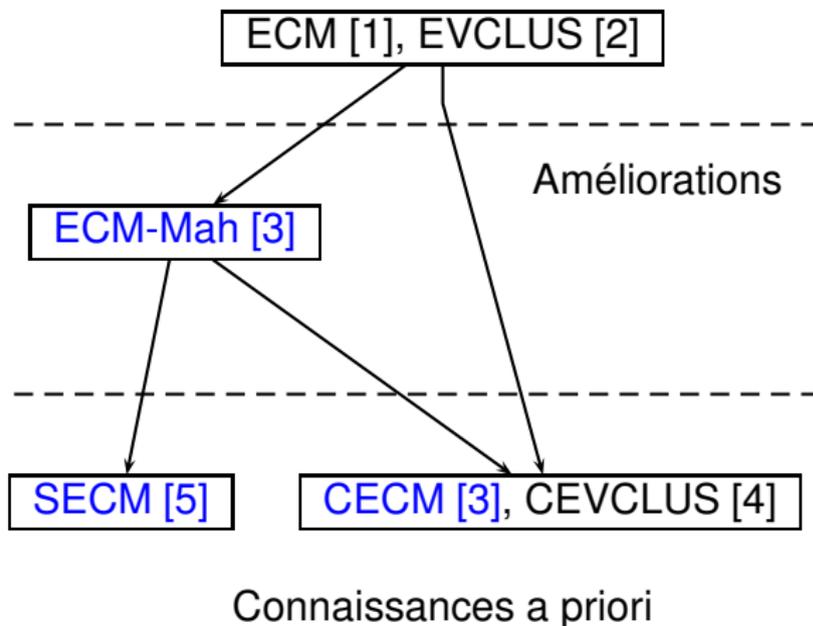
Type de contraintes

Intégrer des contraintes dans la classification automatique :

- au niveau du modèle
 - classes équilibrées
 - information négative : un modèle rejeté
 - au niveau des classes
 - au niveau des objets
-
- étiquettes : $x_i \in \omega_k$
 - étiquettes floues : $x_i \in \{\omega_k, \omega_l\} \Rightarrow x_i \in \omega_k$ OU $x_i \in \omega_l$

Motivations

Classification automatique évidentielle



[1] M.-H. Masson & al, *ECM : An evidential version of the fuzzy c-means algorithm*, 2008



[2] T. Denœux & al, *EVCLUS : evidential clustering of proximity data*, 2004



[3] V. Antoine & al, *CECM : Constrained Evidential C-Means algorithm*, 2012



[4] V. Antoine & al, *CEVCLUS : Evidential clustering with instance-level constraints for relational data*, 2014



[5] V. Antoine & al, *Evidential seed-based semi-supervised clustering*, 2014.

Plan

- Introduction
 - Classification automatique
 - Classification sous contraintes
 - Motivations
- Contributions
 - ECM-Mah
 - SECM
 - CECM
- Expérimentations
 - Comportement des algorithmes
 - Applications
- Conclusion

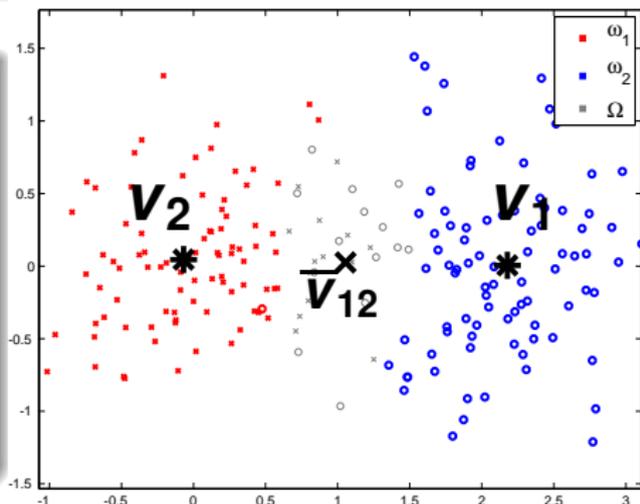
ECM

Principe

- Généralisation des c -moyennes floues
- Objectif : enrichir le concept de partition en utilisant une partition crédale

Modèle géométrique

- Chaque classe ω_k est représentée par un centre v_k
- Centre \bar{v}_j : barycentre des centres associé aux classes composant $A_j \subseteq \Omega$
- Distance Euclidienne d_{ij}^2 entre x_i et \bar{v}_j



ECM

Fonction objectif

$$J_{ECM}(M, V) = \sum_{i=1}^N \sum_{A_j \subseteq \Omega, A_j \neq \emptyset} |A_j|^\alpha m_i(A_j)^\beta d_{ij}^2 + \sum_{i=1}^N \delta^2 m_i(\emptyset)^\beta$$

$$\text{sous contraintes : } \begin{cases} \sum_{A_j \subseteq \Omega, A_j \neq \emptyset} m_i(A_j) + m_i(\emptyset) = 1 \\ m_i(A_j) \geq 0 \quad \forall i, j \end{cases}$$

ECM

Fonction objectif

$$J_{ECM}(M, V) = \sum_{i=1}^N \sum_{\substack{A_j \subseteq \Omega, \\ A_j \neq \emptyset}} |A_j|^\alpha \boxed{m_i(A_j)^\beta d_{ij}^2} + \sum_{i=1}^N \delta^2 m_i(\emptyset)^\beta$$

comme FCM
tous les sous-ensembles

sous contraintes :

$$\begin{cases} \sum_{A_j \subseteq \Omega, A_j \neq \emptyset} m_i(A_j) + m_i(\emptyset) = 1 \\ m_i(A_j) \geq 0 \quad \forall i, j \end{cases}$$

ECM

Fonction objectif

Pénalisation des
sous-ensembles de
fortes cardinalités

Gestion des
points aberrants

$$J_{ECM}(M, V) = \sum_{i=1}^N \sum_{A_j \subseteq \Omega, A_j \neq \emptyset} |A_j|^\alpha m_i(A_j)^\beta d_{ij}^2 + \sum_{i=1}^N \delta^2 m_i(\emptyset)^\beta$$

comme FCM

tous les sous-ensembles

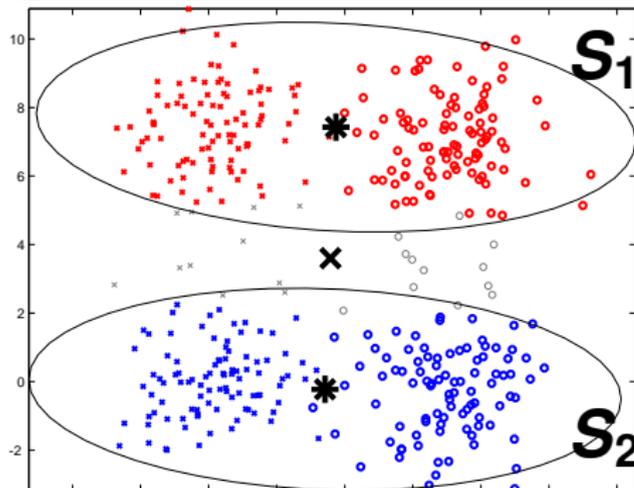
sous contraintes :

$$\begin{cases} \sum_{A_j \subseteq \Omega, A_j \neq \emptyset} m_i(A_j) + m_i(\emptyset) = 1 \\ m_i(A_j) \geq 0 \quad \forall i, j \end{cases}$$

Utilisation d'une métrique adaptative

Distance de Mahalanobis pour chaque classe ω_k

- Chaque classe ω_k est représentée par un centre v_k
- Chaque classe ω_k est associée à une matrice de covariance S_k



Définition

$$d_{ij}^2 = (\mathbf{x}_i - \bar{\mathbf{v}}_j)^t \bar{\mathbf{S}}_j (\mathbf{x}_i - \bar{\mathbf{v}}_j)$$

tel que

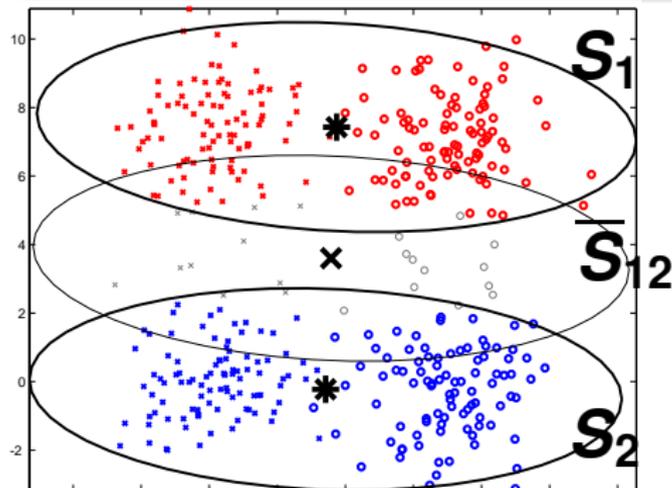
$$\bar{\mathbf{S}}_j = \frac{1}{|A_j|} \sum_{\omega_k \in A_j} \mathbf{S}_k,$$

$$\forall A_j \subseteq \Omega, A_j \neq \emptyset$$

Utilisation d'une métrique adaptative

Distance de Mahalanobis pour chaque classe ω_k

- Chaque classe ω_k est représentée par un centre v_k
- Chaque classe ω_k est associée à une matrice de covariance S_k



Définition

$$d_{ij}^2 = (\mathbf{x}_i - \bar{\mathbf{v}}_j)^t \bar{\mathbf{S}}_j (\mathbf{x}_i - \bar{\mathbf{v}}_j)$$

tel que

$$\bar{\mathbf{S}}_j = \frac{1}{|A_j|} \sum_{\omega_k \in A_j} \mathbf{S}_k,$$

$$\forall A_j \subseteq \Omega, A_j \neq \emptyset$$

Utilisation d'une métrique adaptative

Nouvelle fonction objectif

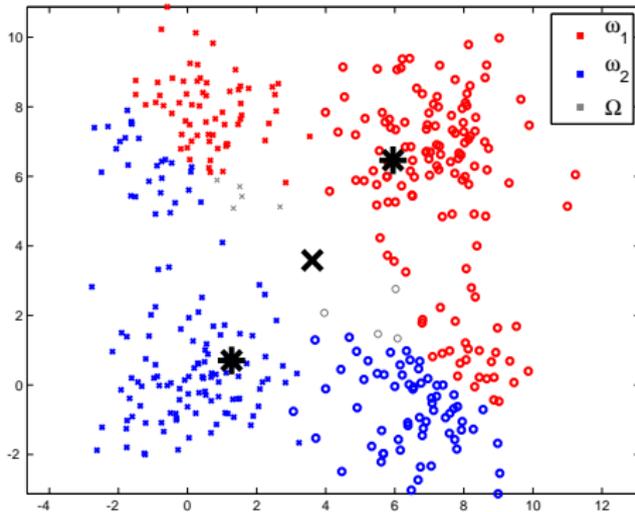
Minimiser J_{ECM} par rapport m_{ij} , \mathbf{v}_k , \mathbf{S}_k s.c. $|\mathbf{S}_k| = 1 \quad \forall k = 1, C$

Optimisation

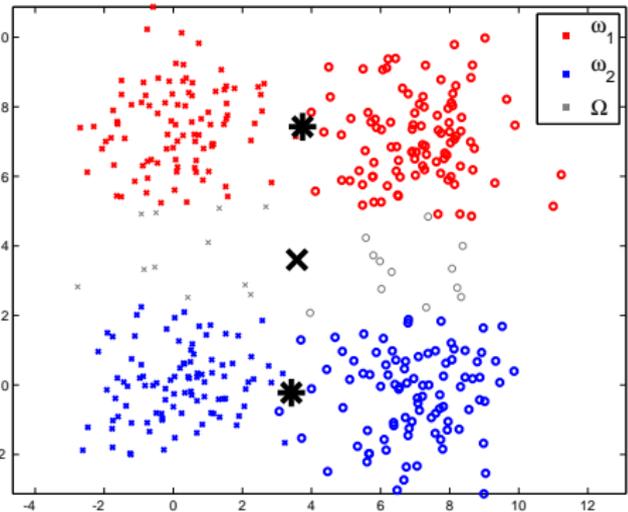
Les conditions de Kuhn–Tucker donnent :

- $m_i(A_j)$ identique à ECM avec une distance Euclidienne
- \mathbf{v}_k : système d'équations linéaires
- \mathbf{S}_k : similaire à Gustafson et Kessel

Utilisation d'une métrique adaptative



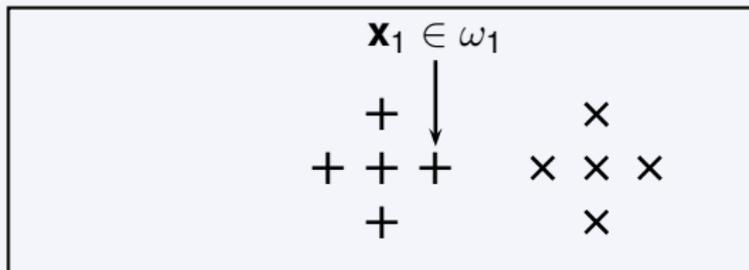
ECM+distance Euclidienne



ECM+distance de Mahalanobis

Ajout d'étiquettes

Exemple

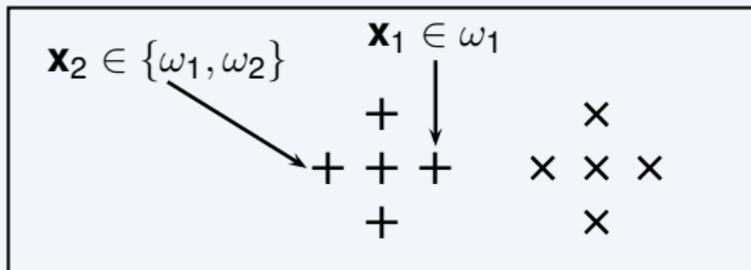


$$\mathbf{x}_1 \in \omega_1 \Rightarrow m_1(\omega_1) = 1$$

A_j	$m_1(A_j)$		
\emptyset	0	0	0
ω_1	1	0	0
ω_2	0	1	0
Ω	0	0	1
	✓	✗	✗

Ajout d'étiquettes

Exemple



$$\mathbf{x}_1 \in \omega_1 \Rightarrow m_1(\omega_1) = 1$$

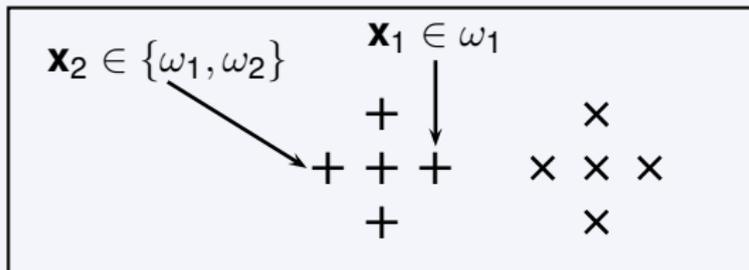
A_j	$m_1(A_j)$		
\emptyset	0	0	0
ω_1	1	0	0
ω_2	0	1	0
Ω	0	0	1
	✓	✗	✗

$$\mathbf{x}_2 \in \{\omega_1, \omega_2\} \Rightarrow m_2(\{\omega_1, \omega_2\}) = 1$$

A_j	$m_2(A_j)$		
\emptyset	0	0	0
ω_1	0	1	0
ω_2	0	0	1
Ω	1	0	0
	✓	✗	✗

Ajout d'étiquettes

Exemple



~~$\mathbf{x}_1 \in \omega_1 \Rightarrow m_1(\omega_1) = 1$~~

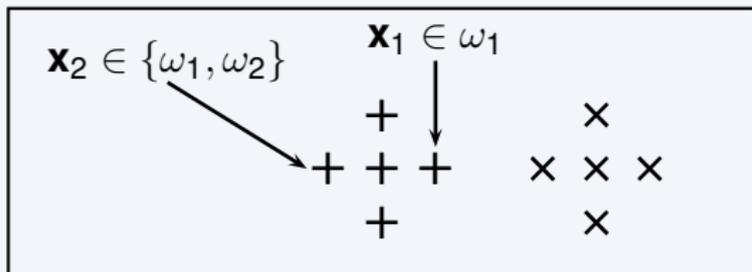
~~$\mathbf{x}_2 \in \{\omega_1, \omega_2\} \Rightarrow m_2(\{\omega_1, \omega_2\}) = 1$~~

~~| A_j | $m_1(A_j)$ | | |
|-------------|------------|---|---|
| \emptyset | 0 | 0 | 0 |
| ω_1 | 1 | 0 | 0 |
| ω_2 | 0 | 1 | 0 |
| Ω | 0 | 0 | 1 |
| | ✓ | ✗ | ✗ |~~
~~| A_j | $m_2(A_j)$ | | |
|-------------|------------|---|---|
| \emptyset | 0 | 0 | 0 |
| ω_1 | 0 | 1 | 0 |
| ω_2 | 0 | 0 | 1 |
| Ω | 1 | 0 | 0 |
| | ✓ | ✗ | ✗ |~~

Trop rigide !

Ajout d'étiquettes

Exemple



$$\mathbf{x}_1 \in \omega_1 \Rightarrow Pl_1(\omega_1) = 1$$

A_j	$m_1(A_j)$		
\emptyset	0	0	0
ω_1	1	0	0
ω_2	0	1	0
Ω	0	0	1
	✓	✗	✓

⇒ robuste au bruit

$$\mathbf{x}_2 \in \{\omega_1, \omega_2\} \Rightarrow Pl_2(\{\omega_1, \omega_2\}) = 1$$

A_j	$m_2(A_j)$			
\emptyset	0	0	0	
ω_1	0	1	0	
ω_2	0	0	1	
Ω	1	0	0	
	✓	✓	✓	

⇒ résultats plus précis

ECM avec des étiquettes

- Idée : si $x_i \in A_j \Rightarrow p_{ij}(A_j)$ élevée

Fonction objectif

$$J_{SECM} = (1 - \varphi) \sum_{i=1}^N \sum_{A_j \subseteq \Omega, A_j \neq \emptyset} |A_j|^\alpha m_i(A_j)^\beta d_{ij}^2 + \sum_{i=1}^N \delta^2 m_i(\emptyset)^\beta \\ + \varphi \sum_{x_i \in A_j} 1 - p_{ij}(A_j),$$

sous contraintes : $\sum m_i(A_j) + m_i(\emptyset) = 1$ and $m_i(A_j) \geq 0 \quad \forall i, j$

ECM avec des étiquettes

- Idée : si $x_j \in A_j \Rightarrow p_{ij}(A_j)$ élevée

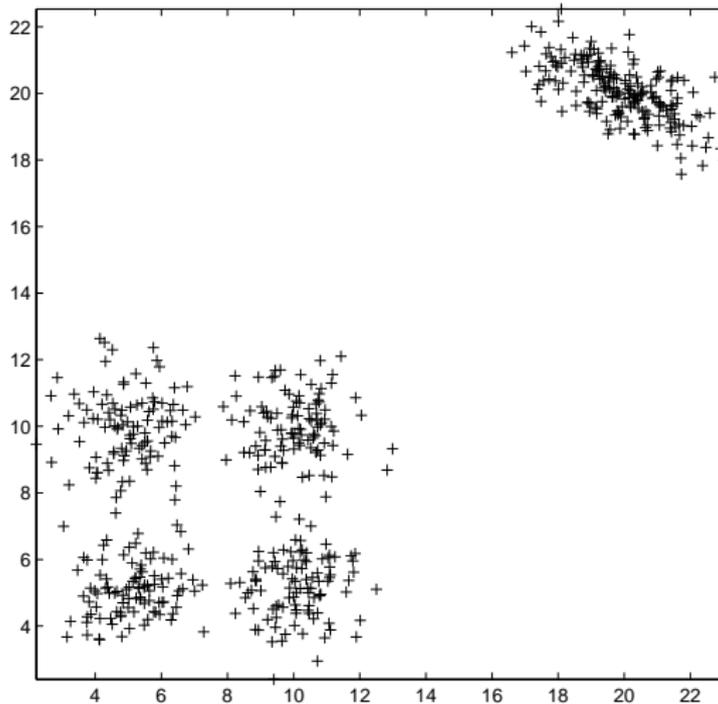
Fonction objectif

$$J_{SECM} = (1 - \varphi) \sum_{i=1}^N \sum_{A_j \subseteq \Omega, A_j \neq \emptyset} |A_j|^\alpha m_i(A_j)^\beta d_{ij}^2 + \sum_{i=1}^N \delta^2 m_i(\emptyset)^\beta + \varphi \sum_{x_j \in A_j} 1 - p_{ij}(A_j),$$

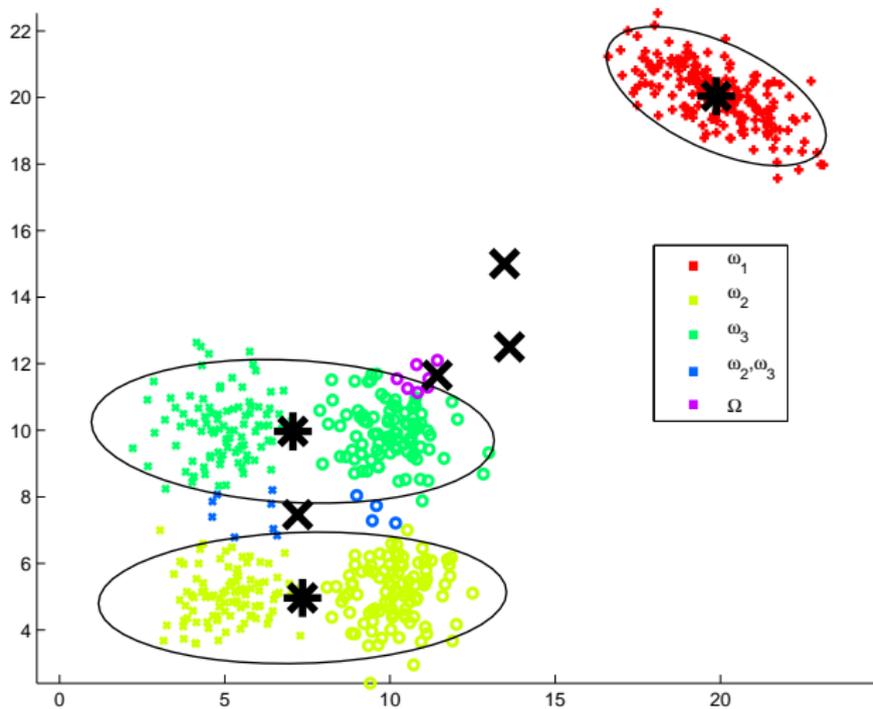
sous contraintes : $\sum m_i(A_j) + m_i(\emptyset) = 1$ and $m_i(A_j) \geq 0 \quad \forall i, j$

\Rightarrow SECM quadratique quand $\beta = 2$, contraintes linéaires

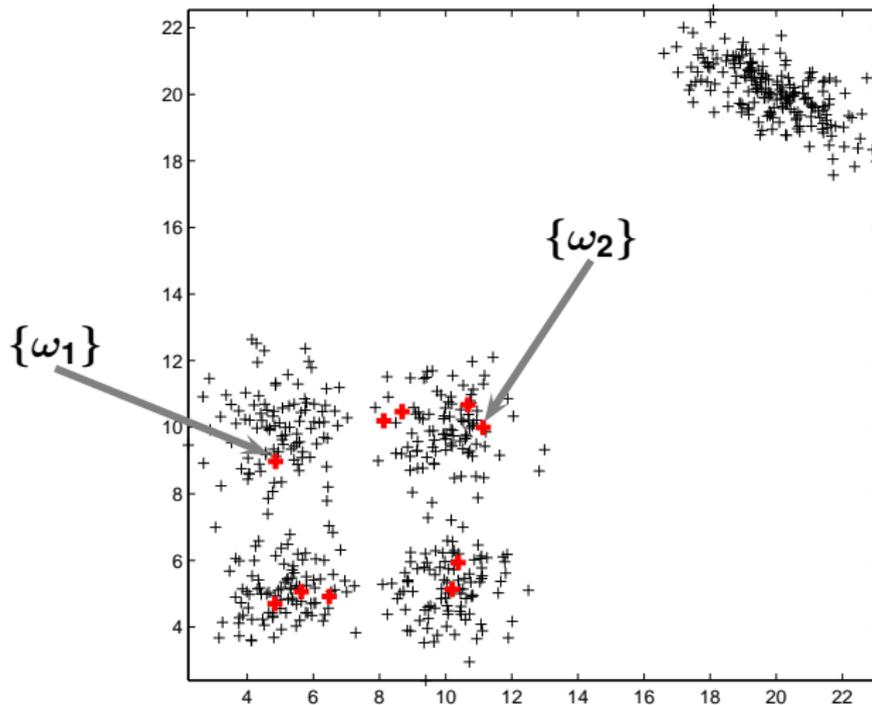
Exemple Toy



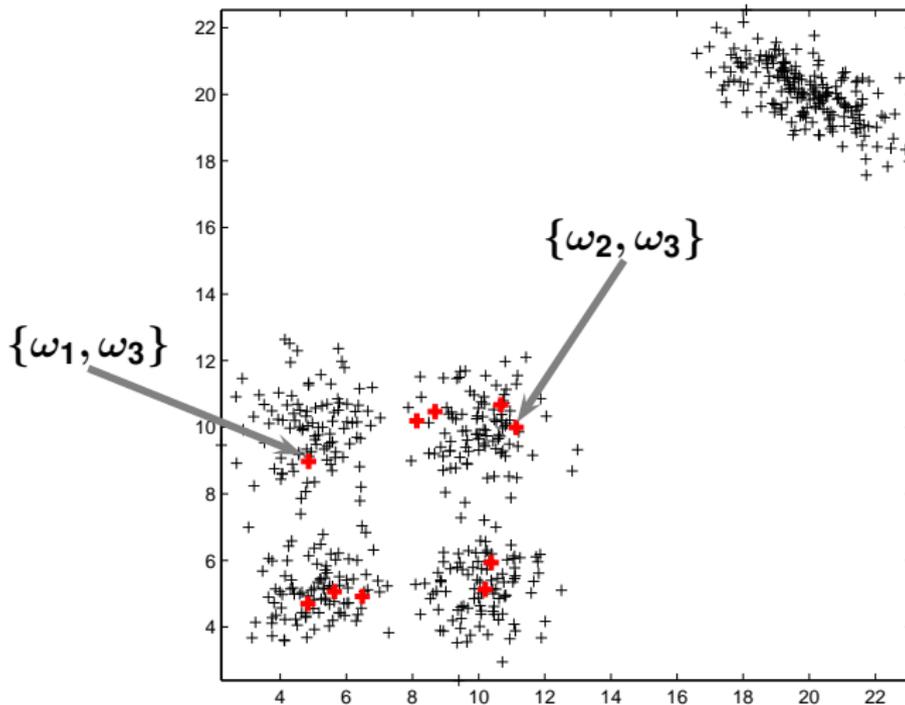
ECM-Mah



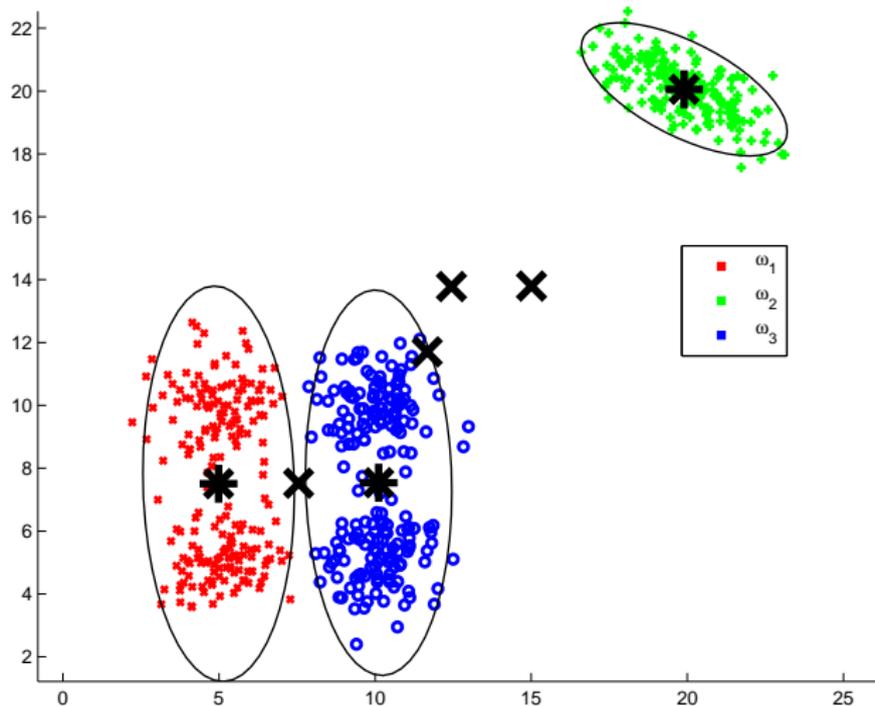
Ajout de contraintes



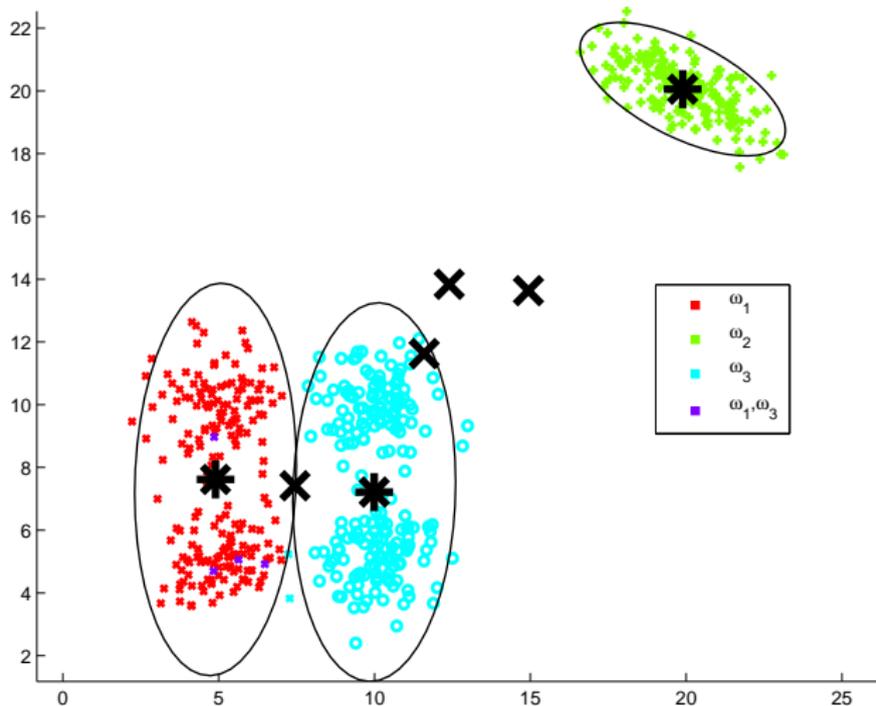
Ajout de contraintes



SECM, contraintes certaines



SECM, contraintes incertaines



Ajout de contraintes ML/CL

Formalisation

- Degré d'appartenance conjointe de $\mathbf{x}_i, \mathbf{x}_j$

$$m_{i \times j}(A \times B) = m_i(A)m_j(B) \quad \forall A, B \subseteq \Omega, A \neq \emptyset, B \neq \emptyset$$

- Dans Ω^2 , les événements

- $\theta \Rightarrow$ " \mathbf{x}_i et \mathbf{x}_j appartiennent à la même classe"
- $\bar{\theta} \Rightarrow$ " \mathbf{x}_i et \mathbf{x}_j sont dans deux classes différentes"

\Rightarrow Plausibilité d'appartenance à la même classe

$$pl_{i \times j}(\theta) = \sum_{A \cap B \neq \emptyset} m_i(A) m_j(B)$$

\Rightarrow Plausibilité d'appartenance à une classe différente

$$pl_{i \times j}(\bar{\theta}) = 1 - m_{i \times j}(\theta) - \sum_{k=1 \dots c} m_i(\{\omega_k\})m_j(\{\omega_k\})$$

Ajout de contraintes ML/CL

Exemple

A	m_1	m_2	m_3	m_4	m_5	\Rightarrow	θ	$p_{l_{1 \times 2}}$	$p_{l_{1 \times 3}}$	$p_{l_{1 \times 4}}$	$p_{l_{1 \times 5}}$
\emptyset	0	0	0	0	1		$\bar{\theta}$	1	0	1	0
ω_1	1	1	0	0	0			0	1	1	0
ω_2	0	0	1	0	0						
Ω	0	0	0	1	0						



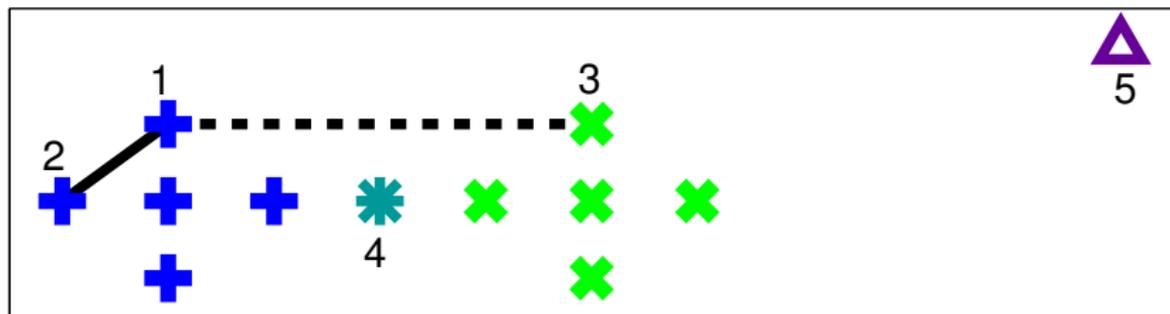
Ajout de contraintes ML/CL

Exemple

A	m_1	m_2	m_3	m_4	m_5
\emptyset	0	0	0	0	1
ω_1	1	1	0	0	0
ω_2	0	0	1	0	0
Ω	0	0	0	1	0

	$pl_{1 \times 2}$	$pl_{1 \times 3}$	$pl_{1 \times 4}$	$pl_{1 \times 5}$
θ	1	0	1	0
$\bar{\theta}$	0	1	1	0

\updownarrow
 $(o_1, o_2) \in \mathcal{M}$ $(o_1, o_3) \in \mathcal{C}$



Ajout de contraintes ML/CL

Exemple

A	m_1	m_2	m_3	m_4	m_5
\emptyset	0	0	0	0	1
ω_1	1	1	0	0	0
ω_2	0	0	1	0	0
Ω	0	0	0	1	0

	$pl_{1 \times 2}$	$pl_{1 \times 3}$	$pl_{1 \times 4}$	$pl_{1 \times 5}$
θ	1	0	1	0
$\bar{\theta}$	0	1	1	0

\updownarrow
 $(o_1, o_2) \in \mathcal{M}$ $(o_1, o_3) \in \mathcal{C}$



Ajout de contraintes : CECM

Principe de base

Si $(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{M} \Rightarrow pl_{i \times j}(\bar{\theta})$ faible et si $(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{C} \Rightarrow pl_{i \times j}(\theta)$ faible

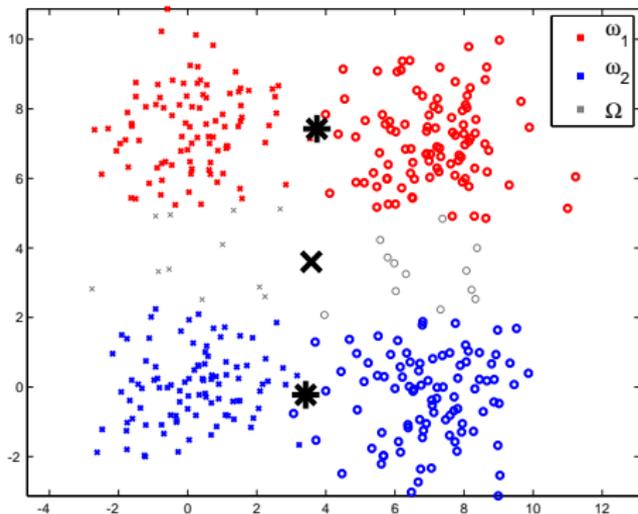
Fonction objectif

$$J_{CECM} = (1 - \phi) \left(\sum_{i=1}^N \sum_{A_j \subseteq \Omega, A_j \neq \emptyset} |A_j|^\alpha m_i(A_j)^\beta d_{ij}^2 + \sum_{i=1}^N \delta^2 m_i(\emptyset)^\beta \right) \\ + \phi \left(\sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{M}} pl_{i \times j}(\bar{\theta}) + \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{C}} pl_{i \times j}(\theta) \right)$$

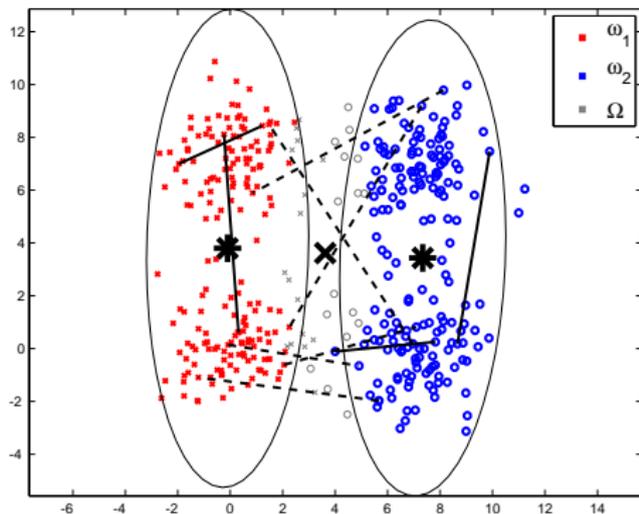
sous contraintes $\sum m_i(A_j) + m_i(\emptyset) = 1$ et $m_i(A_j) \geq 0 \quad \forall i, j$

$\Rightarrow J_{CECM}$ quadratique quand $\beta = 2$, contraintes linéaires

Ajout de contraintes : CECM



ECM+distance de Mahalanobis



CECM+distance de Mahalanobis

Plan

- Introduction
 - Classification automatique
 - Classification sous contraintes
 - Motivations
- Contributions
 - ECM-Mah
 - SECM
 - CECM
- Expérimentations
 - Comportement des algorithmes
 - Applications
- Conclusion

Protocole expérimental

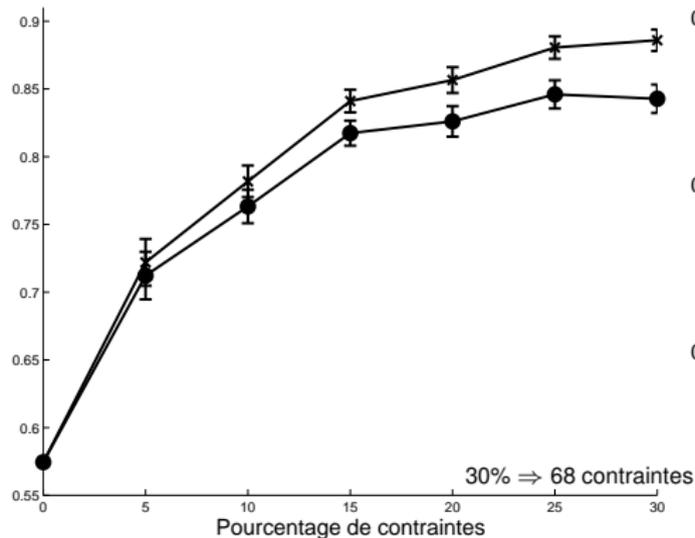
Jeux de données

	# attributs	# objets	# classes
Iris	4	150	3
LettersIJL	16	227	3

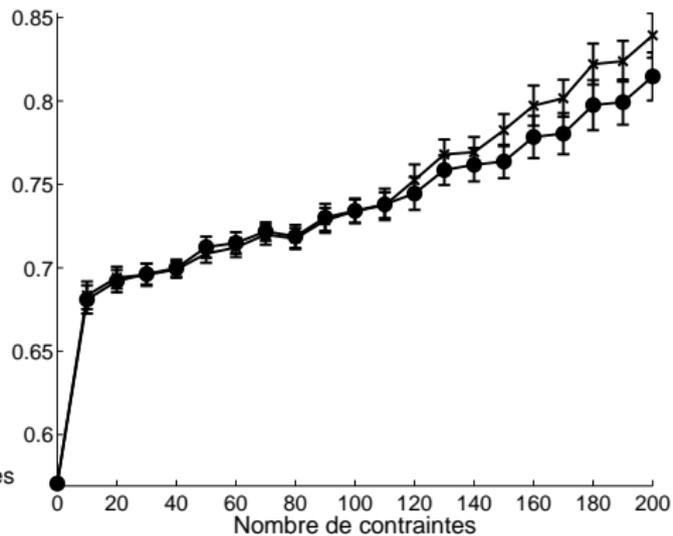
Méthode d'évaluation

- Hyperparamètres
 - SECM, CECM : $\alpha = 1$, $\delta^2 = 1000$ et $\xi = 0.5$
- Sélection aléatoire des contraintes
- Décision : maximum de probabilité pignistique
- Critère : Indice de Rand
 - ⇒ mesure de la concordance globale entre 2 partitions

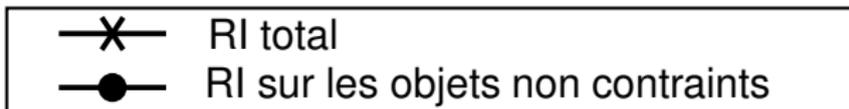
Comportement des algorithmes



LettersIJL, SECM



LettersIJL, CECM



Comparaison avec d'autres algorithmes

	Base algorithmique	respect contraintes	modif. distance
COP [1]	k-Means	Oui	
DML-FCM [2]	k-Means	/	X
CFCM [3]	FCM	/	X
CECM	ECM	/	X



[1] K. Wagstaff & al, *Constrained k-means clustering with background knowledge*, KDID, 2001

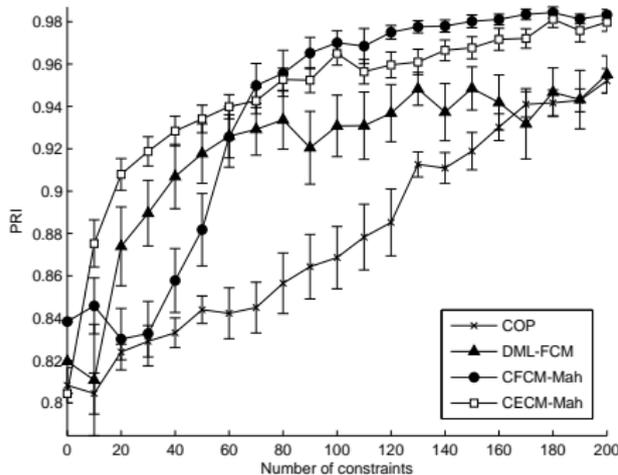


[2] E. Xing & al, *Distance Metric Learning with application to clustering with side-information*, 2002

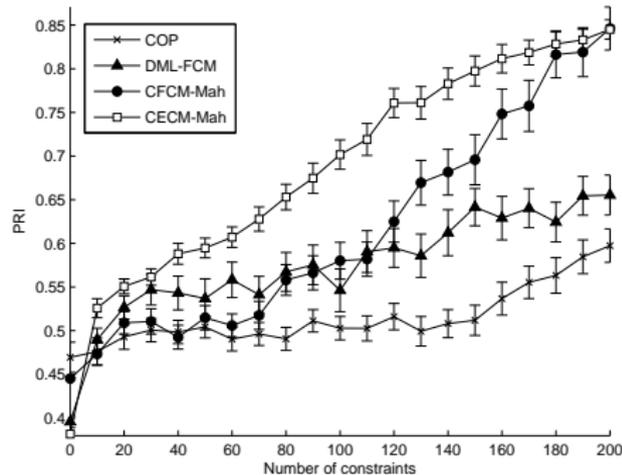


[3] N. Grira & al, *Active semi-supervised fuzzy clustering*, Pattern Recognition, 2008

Comparaison avec d'autres algorithmes



Iris



LettersIJL

Application à la segmentation d'images, SECM

Image avion



image originale



ECM+distance de Mahalanobis

Application à la segmentation d'images, SECM

Image avion



définition des étiquettes



SECM+distance de Mahalanobis

Application à la segmentation d'images, CECM

Image cerveau*

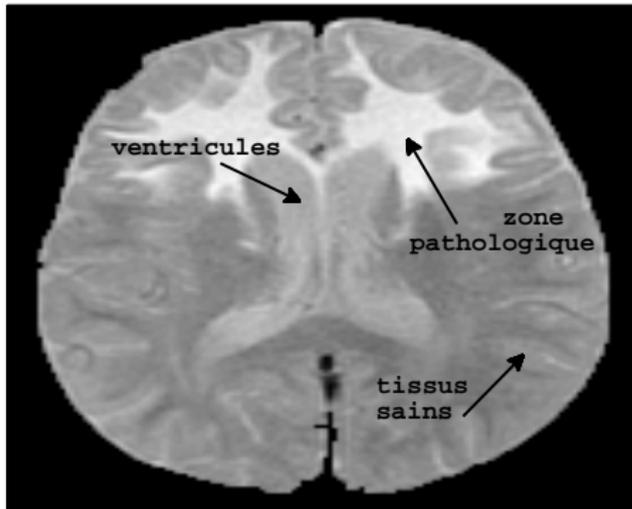
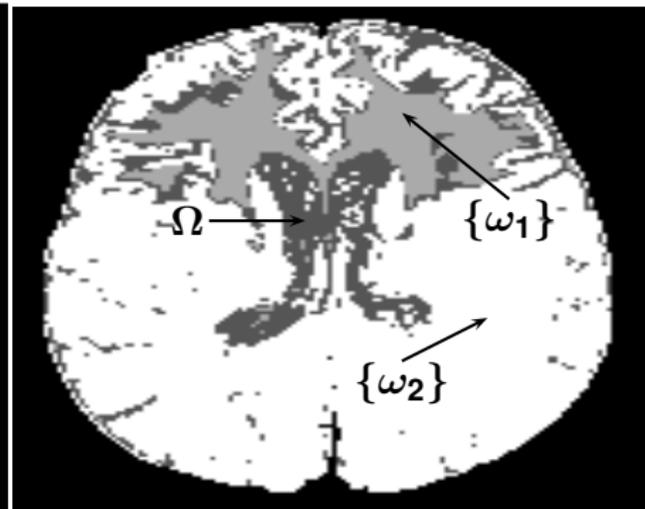


image originale

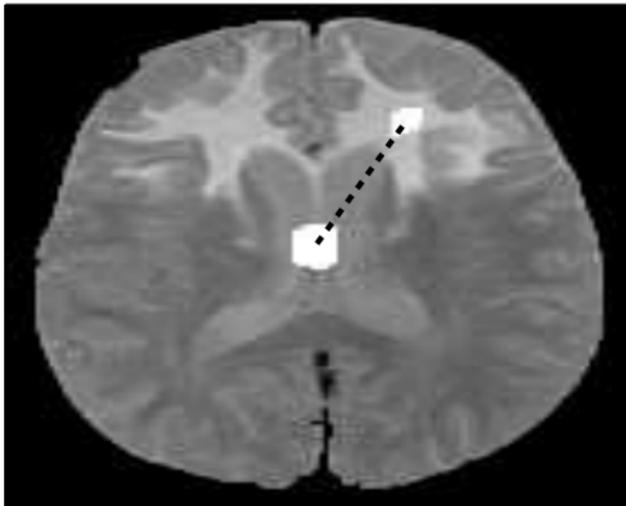


ECM+distance Euclidienne

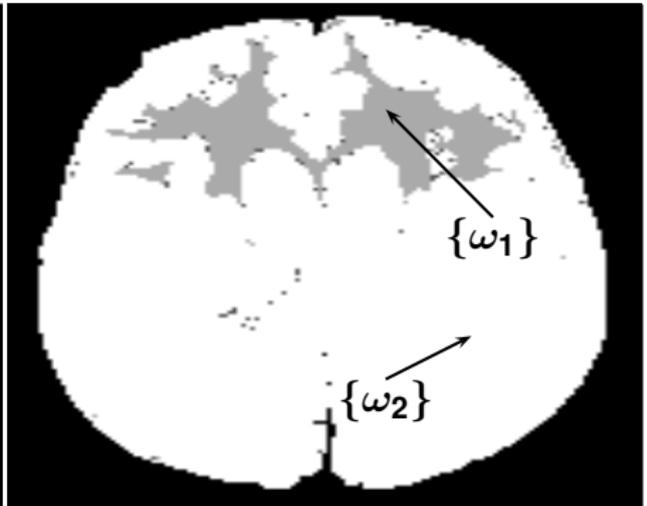
* Fournie par le Prof. Catherine Adamsbaum (Hôpital St Vincent de Paul, Paris, France) et le Prof. Isabelle Bloch (Ecole Nationale Supérieure des Télécommunications, Paris, France)

Application à la segmentation d'images, CECM

Image cerveau*



définition des contraintes



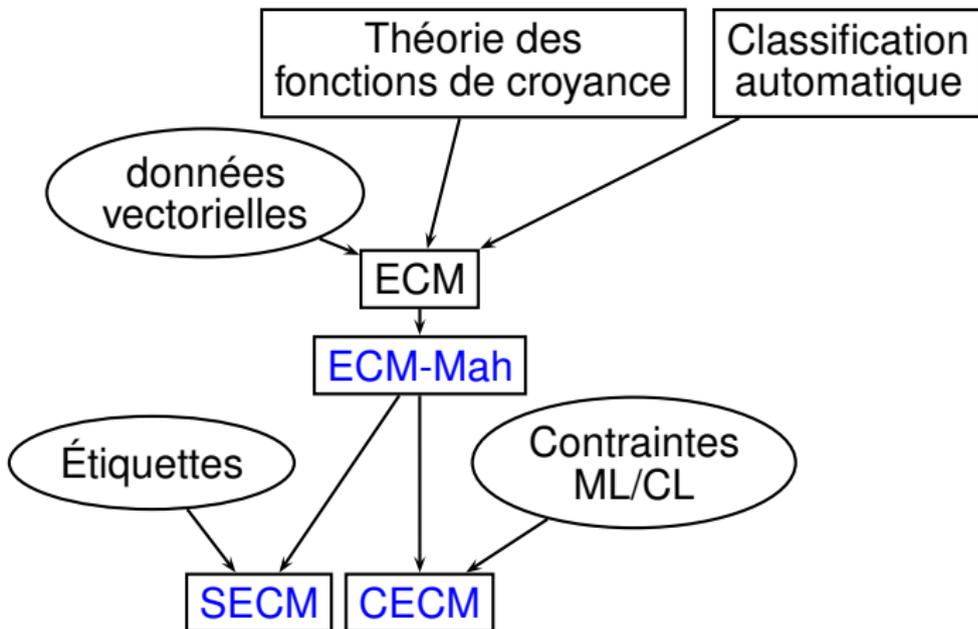
CECM+distance de Mahalanobis

* Fournit par le Prof. Catherine Adamsbaum (Hôpital St Vincent de Paul, Paris, France) et le Prof. Isabelle Bloch (Ecole Nationale Supérieure des Télécommunications, Paris, France)

Plan

- Introduction
 - Classification automatique
 - Classification sous contraintes
 - Motivations
- Contributions
 - ECM-Mah
 - SECM
 - CECM
- Expérimentations
 - Comportement des algorithmes
 - Applications
- Conclusion

Conclusion



Conclusion

Algorithmes de classification sous contraintes

- Les contraintes mènent vers une solution désirée
- Amélioration des performances
- Sensibilité à certains jeux de contraintes

Utilisation d'une partition crédale

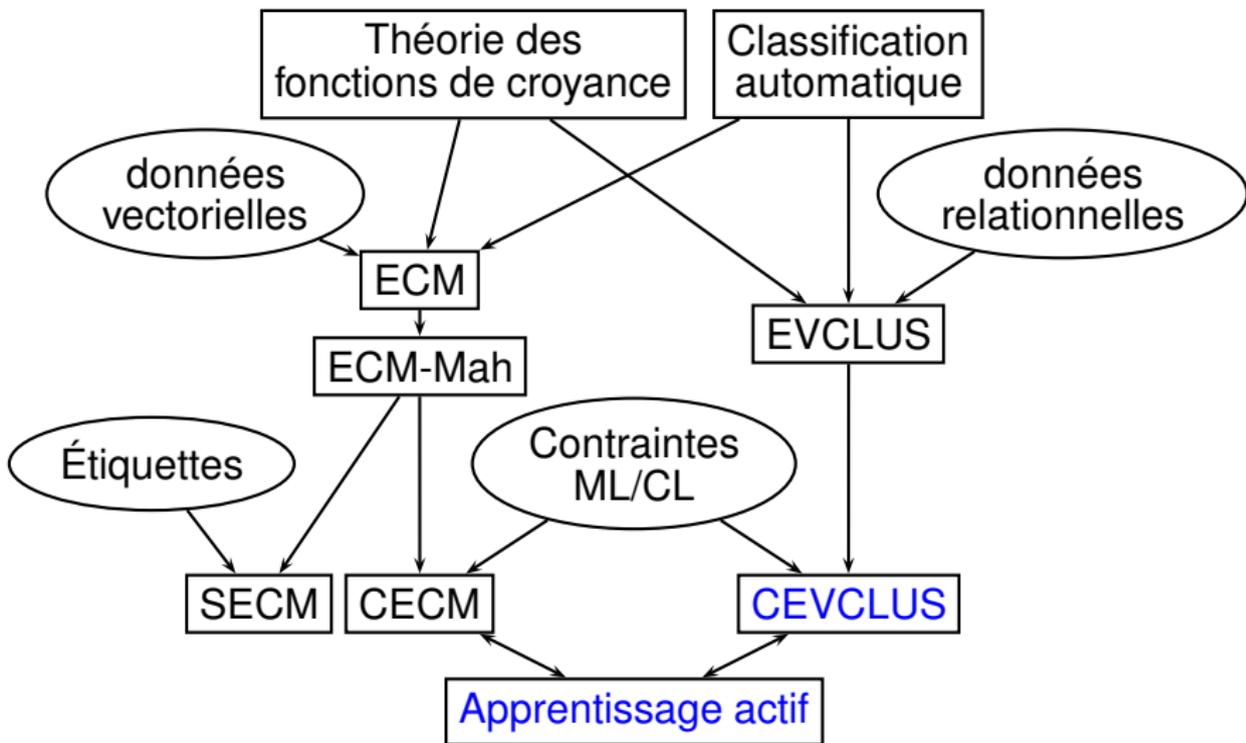
Avantage

- Bon résultats
- Partition riche en information
- Apprentissage actif facilité

Inconvénient

- Complexité calculatoire

Conclusion



Perspectives

Pistes de recherche

- Création de SEVCLUS
- Étude de l'influence des contraintes
 - Apprentissage actif
 - Suppression de contraintes redondantes ou incohérentes
- Amélioration des algorithmes évidentielles
 - diminution de la complexité calculatoire
 - ajout de distances pour ECM

Merci pour votre attention