

Apprentissage actif en classification évidentielle sous contraintes

V. Antoine¹ B. Quost² M.-H. Masson² T. Denœux²

¹Limos, UMR CNRS 6158, Clermont-Ferrand, France

²Heudiasyc, UMR CNRS 6599, Compiègne, France

Séminaire LIP6, février 2014

Plan

- Introduction
 - Classification non supervisée
 - Classification sous contraintes
 - Apprentissage actif
- Contributions
 - Classification évidentiel sous contraintes
 - Apprentissage actif évidentiel
- Expérimentations
 - Comportement des algorithmes
 - Apprentissage actif
- Conclusion

Plan

- Introduction
 - Classification non supervisée
 - Classification sous contraintes
 - Apprentissage actif

- Contributions
 - Classification évidentiel sous contraintes
 - Apprentissage actif évidentiel

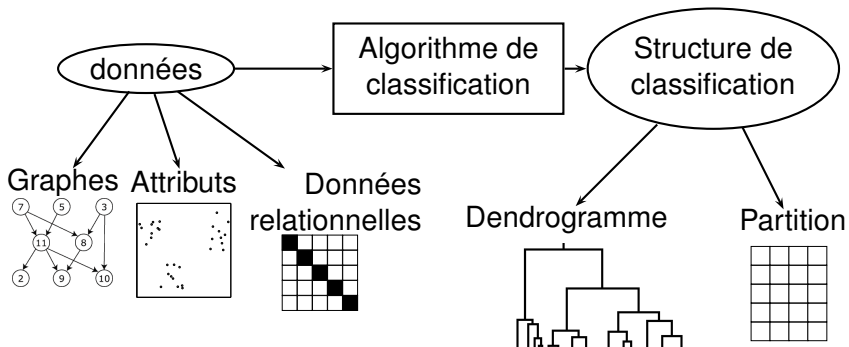
- Expérimentations
 - Comportement des algorithmes
 - Apprentissage actif

- Conclusion

La classification non supervisée

Objectif

Grouper N individus en classes selon une notion de similarité



Partitions

Types de partitions

- Partition dure
 - chaque objet x_i appartient à une classe de manière exclusive
 - exemple d'algorithme : c -moyennes
- Partition floue
 - définition d'un degré d'appartenance de chaque objet x_i à chaque classe ω_k : u_{ik}
 - exemple d'algorithme : FCM
- Partition crédale
 - définition d'un degré de croyance de chaque objet x_i à chaque sous-ensemble A_j de $\Omega = \{\omega_1, \dots, \omega_c\}$: m_{ij}
 - exemples d'algorithmes : ECM, EVCLUS

Théorie des fonctions de croyance

Représentation de l'information

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

Fonction de masse m

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

Fonction de plausibilité

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

Prise de décision

Transformation pignistique

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

Non-spécificité

$$N(m) = \sum_{A \subseteq \Omega \setminus \emptyset} m(A) \log_2 |A| + m(\emptyset) \log_2 |\Omega|$$

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 : **RECM, EVCLUS**

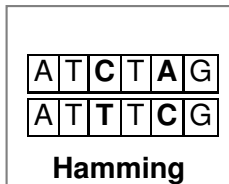
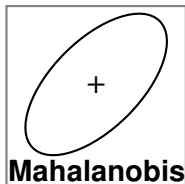
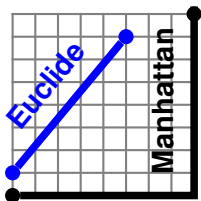
La classification non supervisée

Définition

Grouper N individus en classes selon une notion de similarité

Problématiques sous-jacentes

- Comment définir la notion de similarité ?
- Quelle partition finale choisir ?
- Les classes sont-elles équilibrées ?



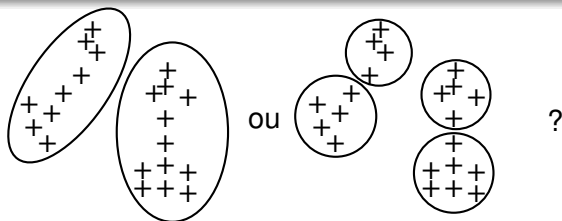
La classification non supervisée

Définition

Grouper N individus en classes selon une notion de similarité

Problématiques sous-jacentes

- Comment définir la notion de similarité ?
- Quelle partition finale choisir ?
- Les classes sont-elles équilibrées ?



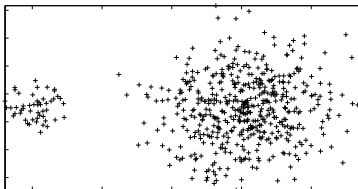
La classification non supervisée

Définition

Grouper N individus en classes selon une notion de similarité

Problématiques sous-jacentes

- Comment définir la notion de similarité ?
- Quelle partition finale choisir ?
- Les classes sont-elles équilibrées ?

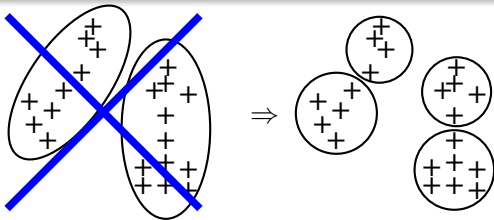


Classification sous contraintes

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

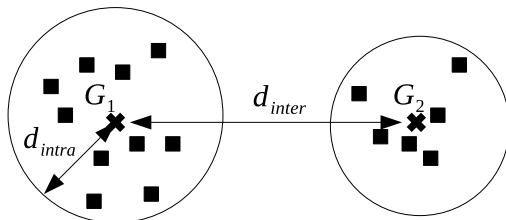


Classification sous contraintes

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

Contraintes Must-Link / Cannot-Link

Types d'algorithmes

- Respect total des contraintes [1]
 - ⇒ Limite l'exploration des solutions MAIS problème de faisabilité
- Respect partiel des contraintes [2, 3]

Méthodes d'intégration de contraintes

Ajout de contraintes dans des algorithmes de base (FCM, CL)

- Ajout d'une pénalité dans la fonction objectif [3]
- Modification des distances [2]
 - ⇒ Performant pour des classes de formes particulières
 - ⇒ Sensible aux contraintes choisies



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



[2] D. Klein & al, *From Instance-level Constraints to Space-level Constraints : Making the Most of Prior Knowledge in Data Clustering*, ICML, 2002

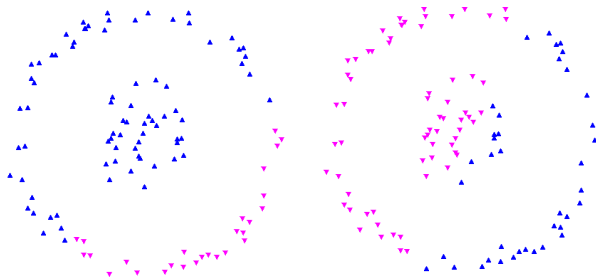


[3] N. Grira & al, *Active semi-supervised fuzzy clustering*, Pattern Recognition, (41)-5 p1834-1844, 2008

Classification sous contraintes : comportement

Exemple 1 : Données avec des centres de gravité confondus

<http://nlp.stanford.edu/danklein/demos/constrained-clustering-demo.shtml>



CCL [2]

COP [1]



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

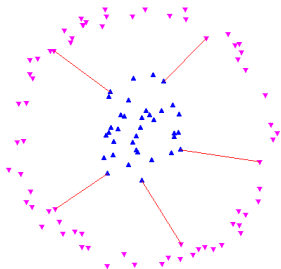


[2] D. Klein & al, *From Instance-level Constraints to Space-level Constraints : Making the Most of Prior Knowledge in Data Clustering*, ICML, 2002

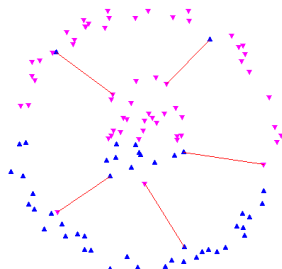
Classification sous contraintes : comportement

Exemple 1 : Données avec des centres de gravité confondus

<http://nlp.stanford.edu/danklein/demos/constrained-clustering-demo.shtml>



CCL [2]



COP [1]



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

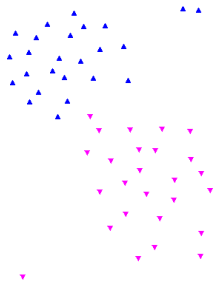


[2] D. Klein & al, *From Instance-level Constraints to Space-level Constraints : Making the Most of Prior Knowledge in Data Clustering*, ICML, 2002

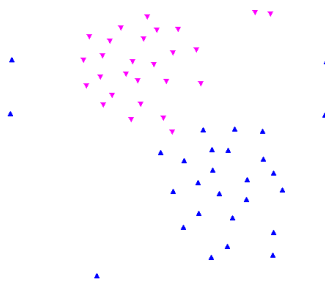
Classification sous contraintes : comportement

Exemple 2 : Données avec des centres de gravité distincts

<http://nlp.stanford.edu/danklein/demos/constrained-clustering-demo.shtml>



SpatialCL [2]



COP-k-Means [1]



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

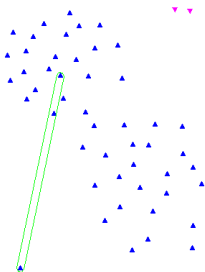


[2] D. Klein & al, *From Instance-level Constraints to Space-level Constraints : Making the Most of Prior Knowledge in Data Clustering*, ICML, 2002

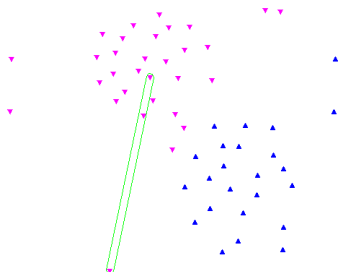
Classification sous contraintes : comportement

Exemple 2 : Données avec des centres de gravité distincts

<http://nlp.stanford.edu/danklein/demos/constrained-clustering-demo.shtml>



SpatialCL [2]



COP-k-Means [1]



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



[2] D. Klein & al, *From Instance-level Constraints to Space-level Constraints : Making the Most of Prior Knowledge in Data Clustering*, ICML, 2002

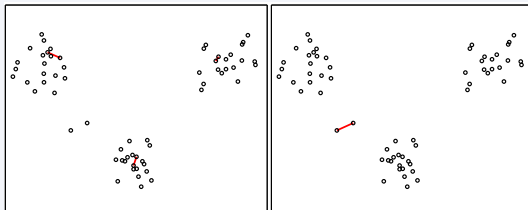
Classification sous contraintes

Problématique

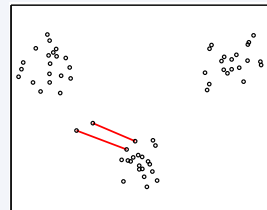
Certaines contraintes dégradent la solution [1]

⇒ sélection intelligente des contraintes

Contraintes non informative



Redondance



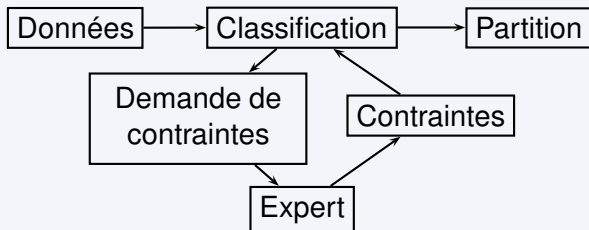
[1] K. Wagstaff & al, *Value, cost, and sharing : Open issues in constrained clustering*, KDID, 2007

Apprentissage actif

Motivations

- Obtenir des contraintes intéressantes à bas coût
- Améliorer des résultats de classification

Principe de l'apprentissage actif



Plan

- Introduction
 - Classification non supervisée
 - Classification sous contraintes
 - Apprentissage actif
- Contributions
 - Classification évidentiel sous contraintes
 - Apprentissage actif évidentiel
- Expérimentations
 - Comportement des algorithmes
 - Apprentissage actif
- Conclusion

Ajout de contraintes

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

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

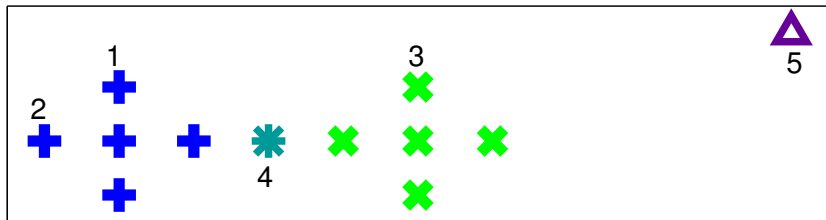
\Rightarrow Plausibilité d'appartenance à une classe différente

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

Ajout de contraintes

Exemple

A	m_1	m_2	m_3	m_4	m_5	\Rightarrow	θ	$p_{1 \times 2}$	$p_{1 \times 3}$	$p_{1 \times 4}$	$p_{1 \times 5}$
\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						



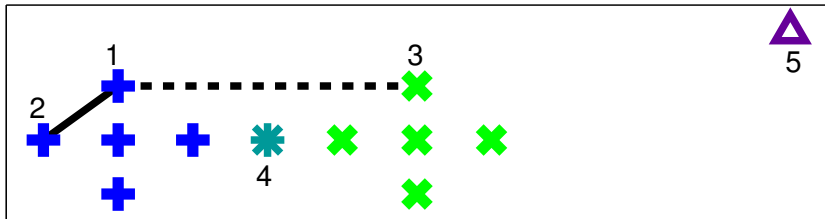
Ajout de contraintes

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



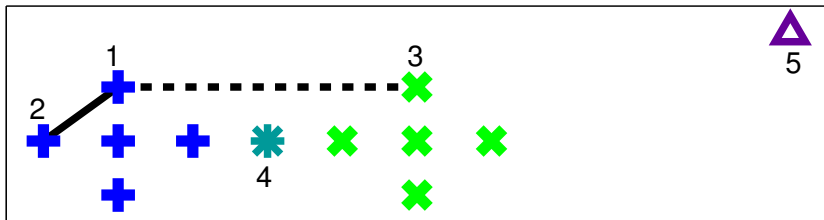
Ajout de contraintes

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

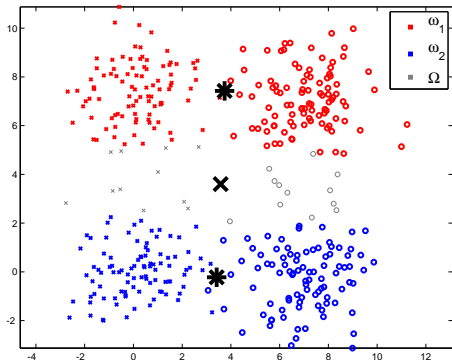
Fonction objectif

$$J_{CECM} = J_{ECM} + \xi \left(\sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{M}} pl_{i \times j}(\bar{\theta}) + \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{C}} pl_{i \times j}(\theta) \right)$$

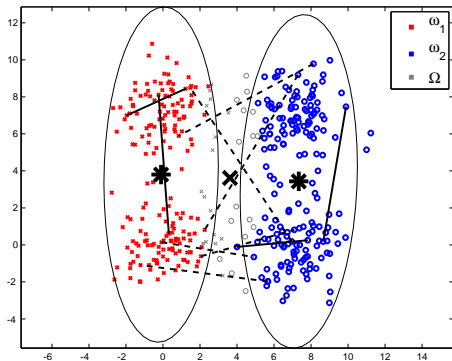
sous contraintes $\sum m_i(A_j) + m_i(\emptyset) = 1$ et $m_i(A_j) \geq 0 \quad \forall i, j$

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

Ajout de contraintes : CECM



ECM+distance de Mahalanobis



CECM+distance de Mahalanobis

Ajout de contraintes : CEVCLUS

Formulation mathématique

$$\text{Cas d'un Must-Link} \quad : \quad J_{ML} = \sum_{(o_i, o_j) \in \mathcal{M}} pl_{i \times j}(\bar{\theta}) + 1 - pl_{i \times j}(\theta)$$

$$\text{Cas d'un Cannot-Link} \quad : \quad J_{CL} = \sum_{(o_i, o_j) \in \mathcal{C}} pl_{i \times j}(\theta) + 1 - pl_{i \times j}(\bar{\theta})$$

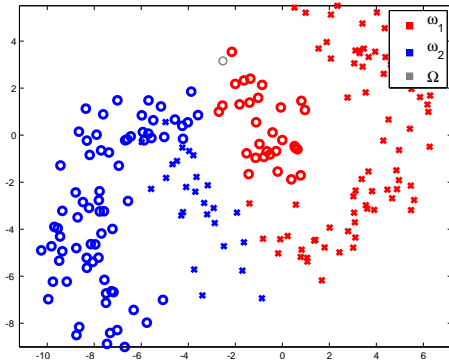
Fonction objectif

$$J_{CEVCLUS} = J_{EVCLUS} + \xi \frac{1}{2(|\mathcal{M}| + |\mathcal{C}|)} (J_{ML} + J_{CL})$$

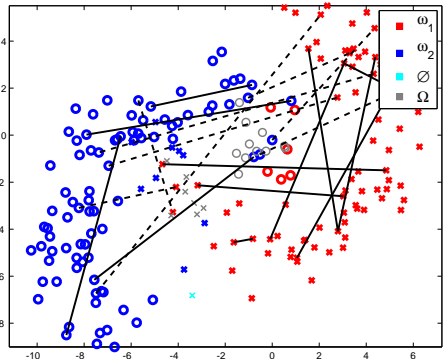
$$\text{sous contraintes} \quad \sum m_i(A_j) + m_i(\emptyset) = 1 \quad \text{et} \quad m_i(A_j) \geq 0 \quad \forall i, j$$

⇒ Minimisation par une méthode de descente de gradient

Jeu de données Banana



EVCLUS



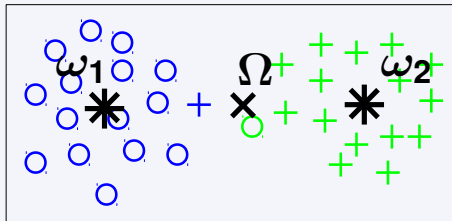
CEVCLUS

Apprentissage actif évidentiel

Principe

- Sélection automatique de paires d'objets
 - Point dont la classe est incertaine
 - Point dont la classe est certaine
 ⇒ mesure de non spécificité

Règle

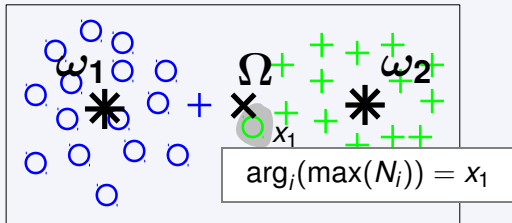


Apprentissage actif évidentiel

Principe

- Sélection automatique de paires d'objets
 - Point dont la classe est incertaine
 - Point dont la classe est certaine
 ⇒ mesure de non spécificité

Règle

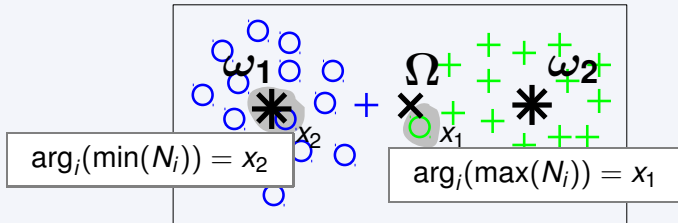


Apprentissage actif évidentiel

Principe

- Sélection automatique de paires d'objets
 - Point dont la classe est incertaine
 - Point dont la classe est certaine
 ⇒ mesure de non spécificité

Règle



Apprentissage actif évidentiel

Principe

- Sélection automatique de paires d'objets
 - Point dont la classe est incertaine
 - Point dont la classe est certaine
- ⇒ mesure de non spécificité

Règle

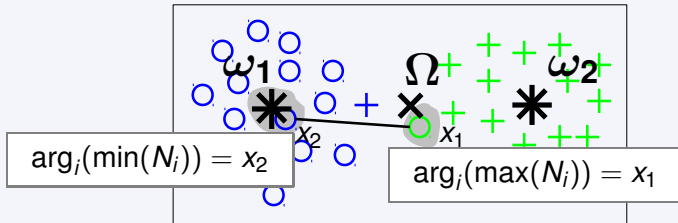
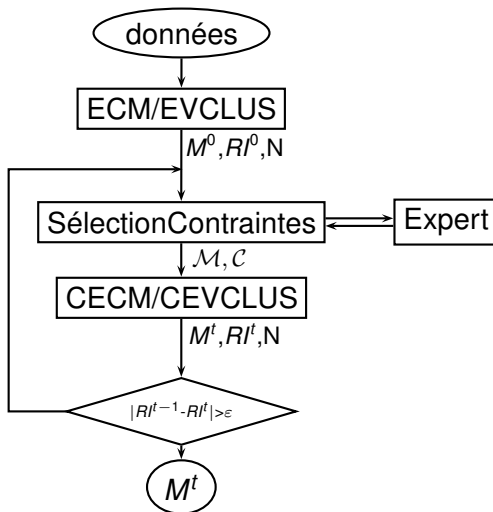


Schéma global



Plan

- Introduction
 - Classification non supervisée
 - Classification sous contraintes
 - Apprentissage actif
- Contributions
 - Classification évidentiel sous contraintes
 - Apprentissage actif évidentiel
- Expérimentations
 - Comportement des algorithmes
 - Apprentissage actif
- Conclusion

Protocole expérimental

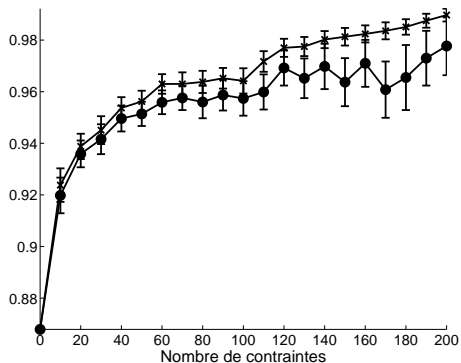
Jeux de données

	# attributs	# objets	# classes
iris	4	150	3
Wine	13	178	3

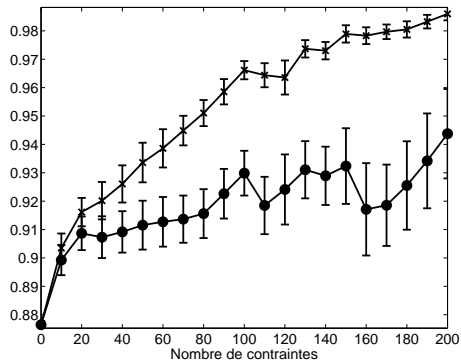
Méthode d'évaluation

- Contraintes
 - ⇒ Sélection aléatoire
- Décision
 - ⇒ Maximum de probabilité pignistique
- Critère : Indice de Rand
 - ⇒ mesure de la concordance globale entre 2 partitions

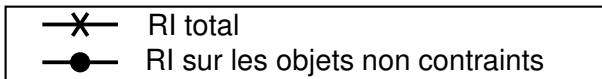
Comportement des algorithmes



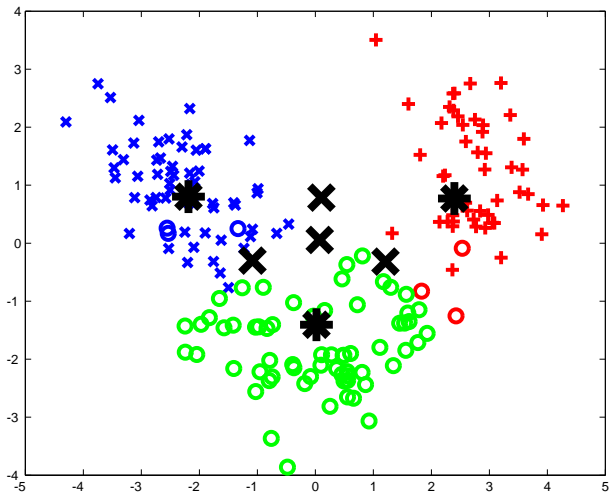
Iris, CECM



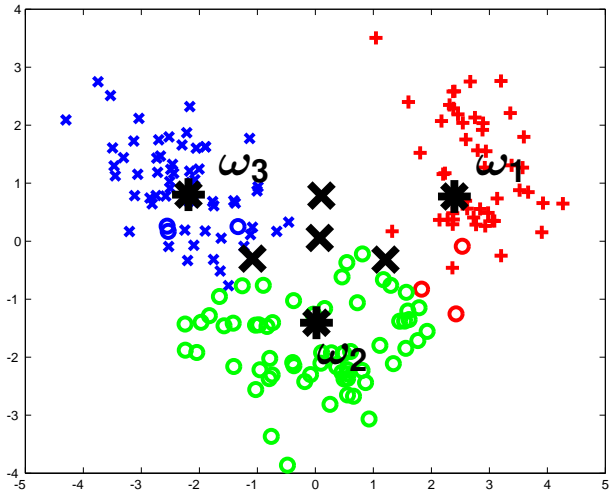
Iris, CEVCLUS



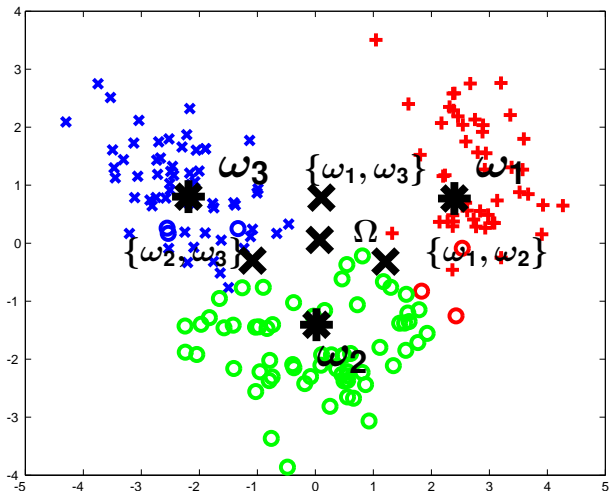
Contraintes mal choisies



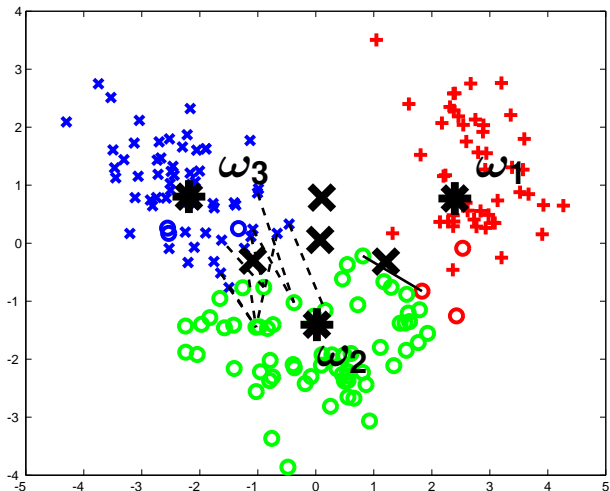
Contraintes mal choisies



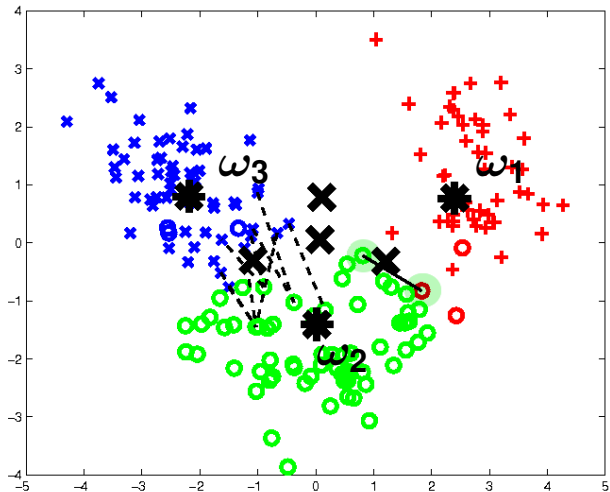
Contraintes mal choisies



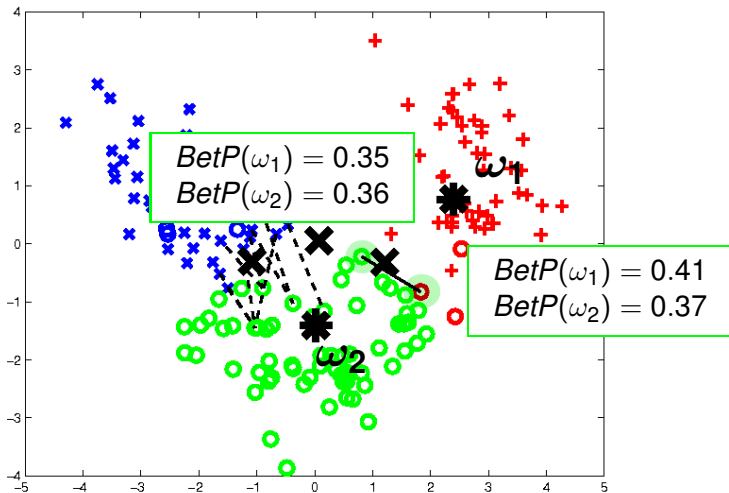
Contraintes mal choisies



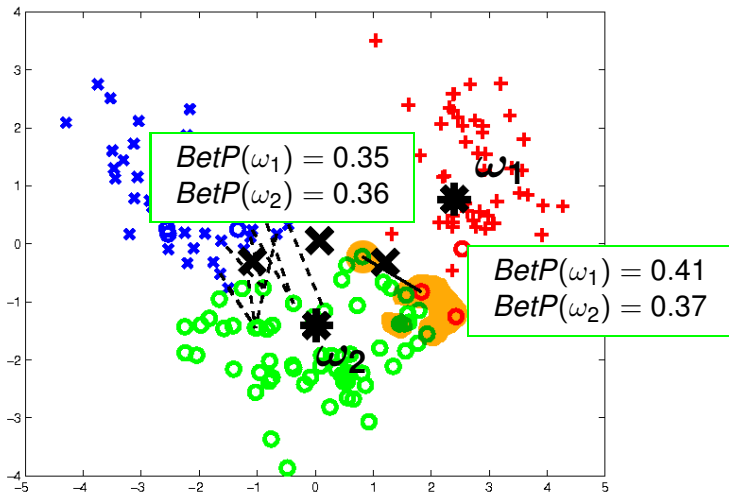
Contraintes mal choisies



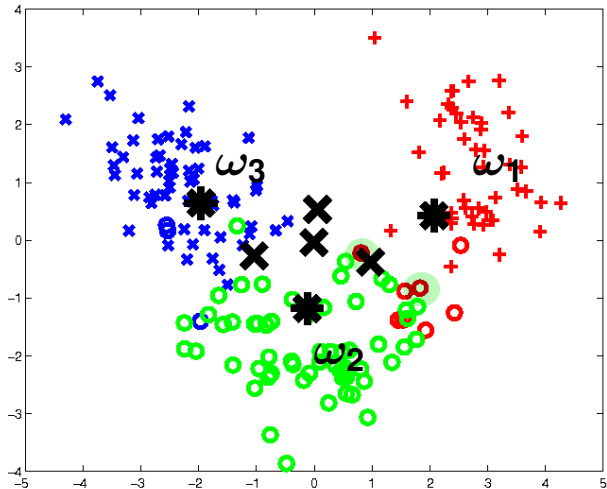
Contraintes mal choisies



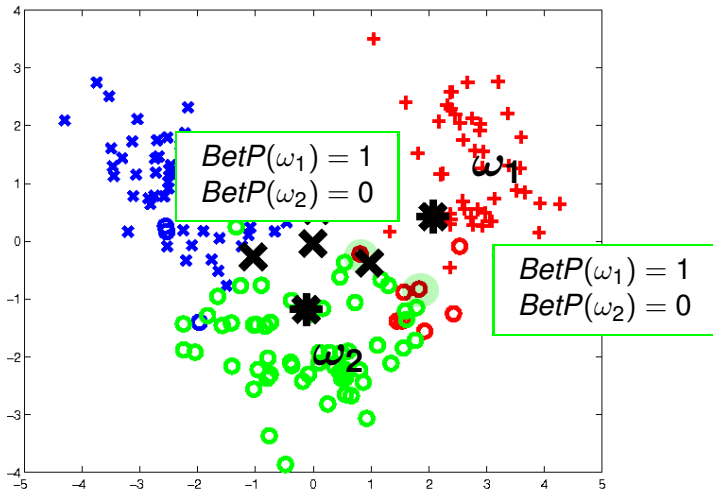
Contraintes mal choisies



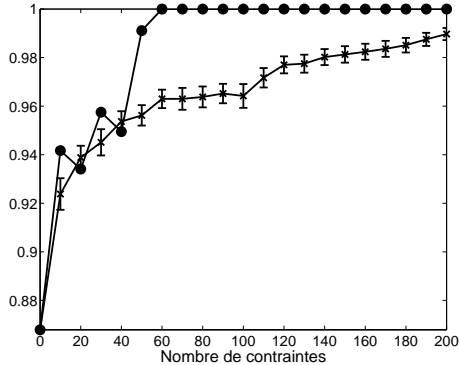
Contraintes mal choisies



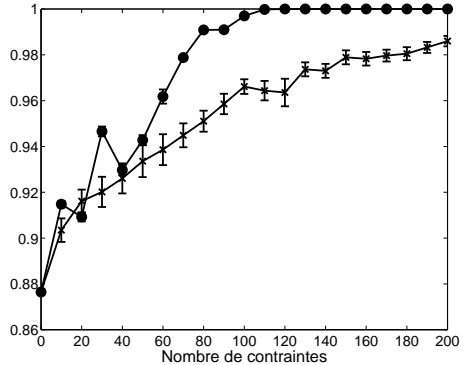
Contraintes mal choisies



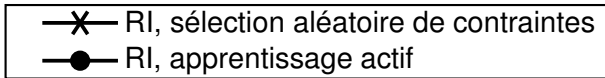
Apprentissage actif



Iris, CECM



Iris, CEVCLUS



Application pour CEVCLUS

Données ChickenPieces

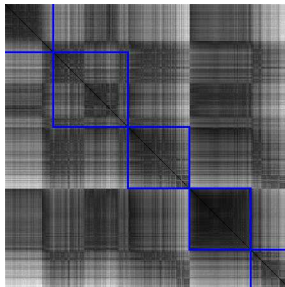
- 446 objets répartis dans 5 classes

Données images ⇒



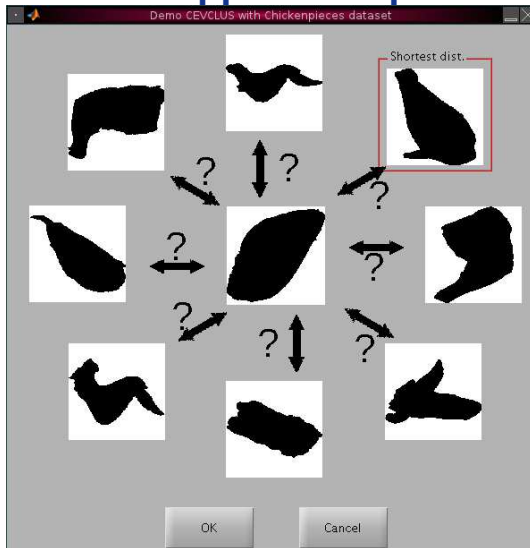
...

⇒ Matrice de dissimilarité



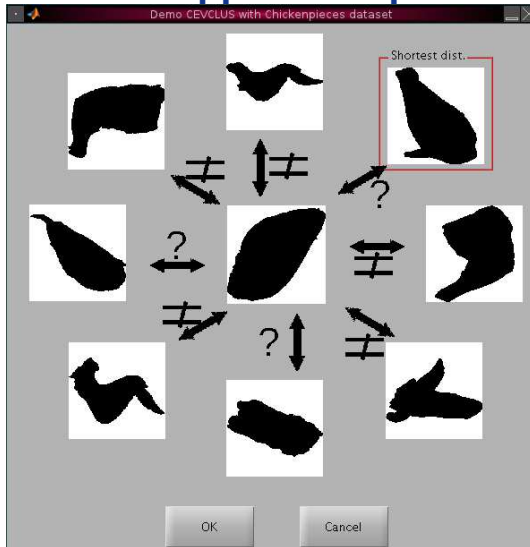
⇒

Application pour CEVCLUS



Ité	#cont.	RI
0	0	0.76

Application pour CEVCLUS



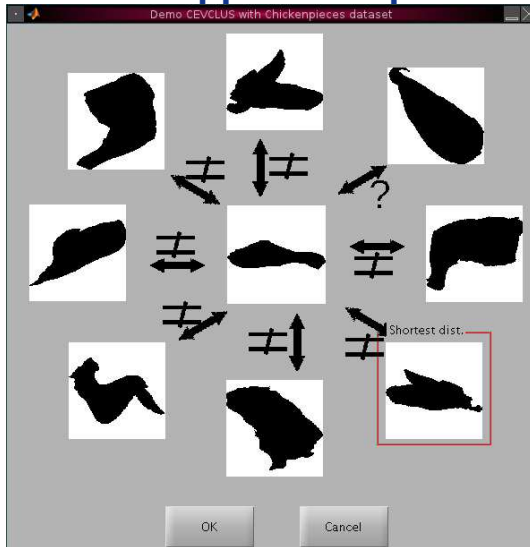
Ité	#cont.	RI
0	0	0.76
1	5	0.80

Application pour CEVCLUS



Ité	#cont.	RI
0	0	0.76
1	5	0.80
2	12	0.82

Application pour CEVCLUS

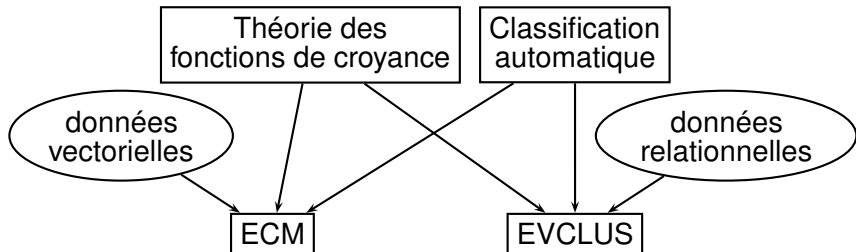


Ité	#cont.	RI
0	0	0.76
1	5	0.80
2	12	0.82
3	19	0.83

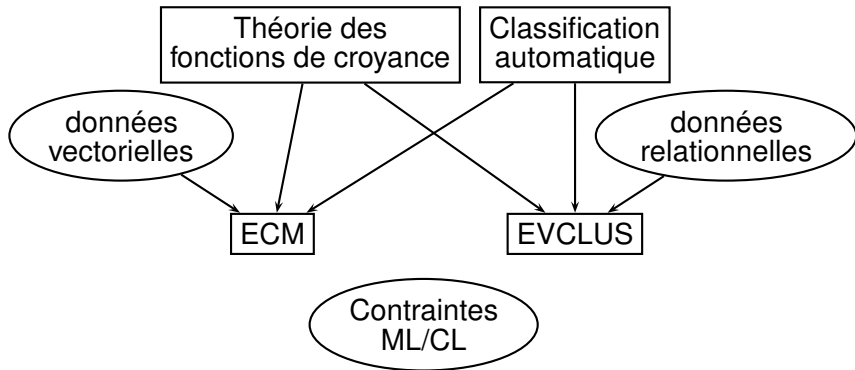
Plan

- Introduction
 - Classification non supervisée
 - Classification sous contraintes
 - Apprentissage actif
- Contributions
 - Classification évidentiel sous contraintes
 - Apprentissage actif évidentiel
- Expérimentations
 - Comportement des algorithmes
 - Apprentissage actif
- Conclusion

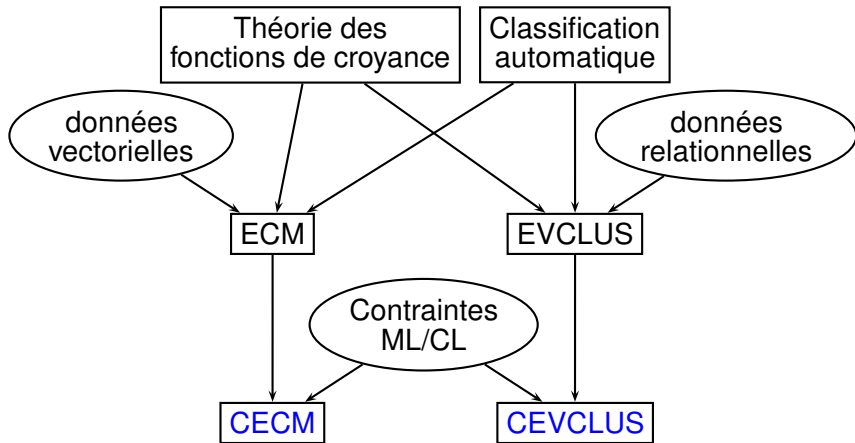
Conclusion



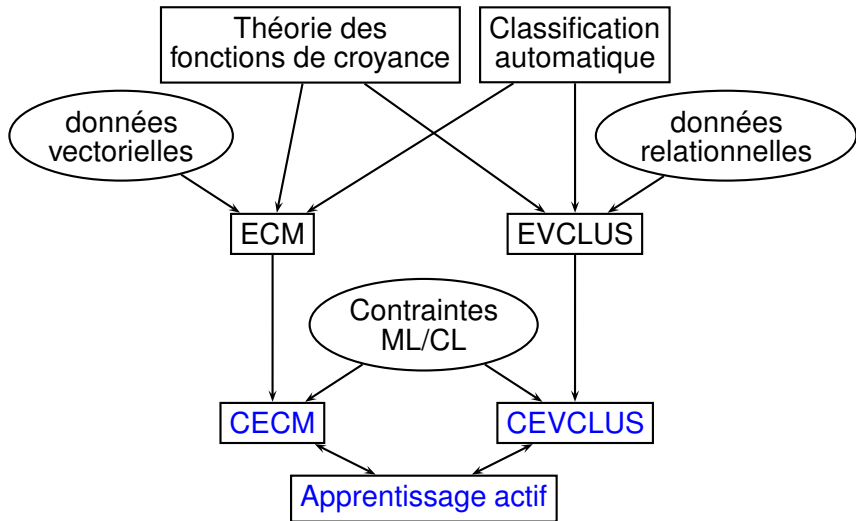
Conclusion



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

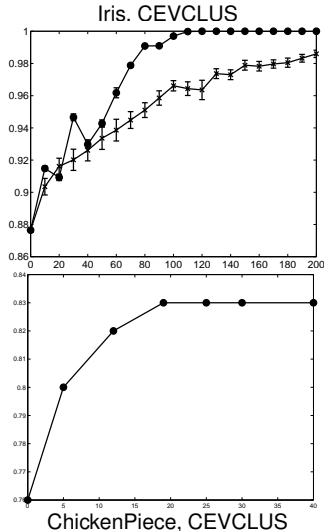
Avantage

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

Inconvénient

- Complexité calculatoire

Pespectives



Apprentissage actif

- Chute des performances
 - ⇒ Phase d'exploration
 - ⇒ Phase d'exploitation ?
- Blocage sur un minimum local
 - ⇒ gestion particulière de \emptyset ?
 - ⇒ autres mesures que N ?

Pistes de recherche

- Étude de l'influence des contraintes
 - Selon les jeux de données
 - Selon les algorithmes de classification sous contraintes
 - ⇒ Mesures d'utilité
 - ⇒ Apprentissage actif plus robuste
 - ⇒ Suppression de contraintes redondantes ou incohérentes
- Utilisation des contraintes en prétraitement
 - Choix d'une métrique adaptée pour CEVCLUS
 - Initialisation des centres de gravité pour CECM

Merci pour votre attention