Evidential clustering with label constraints [1]

V. Antoine

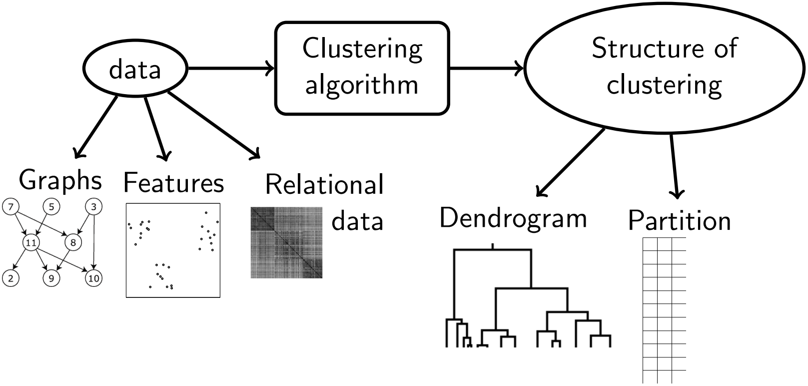
Clermont-Auvergne University, LIMOS, UMR CNRS 6158, France

July 2023

[1] V. Antoine & al, *Fast semi-supervised evidential clustering*, IJAR, 2021

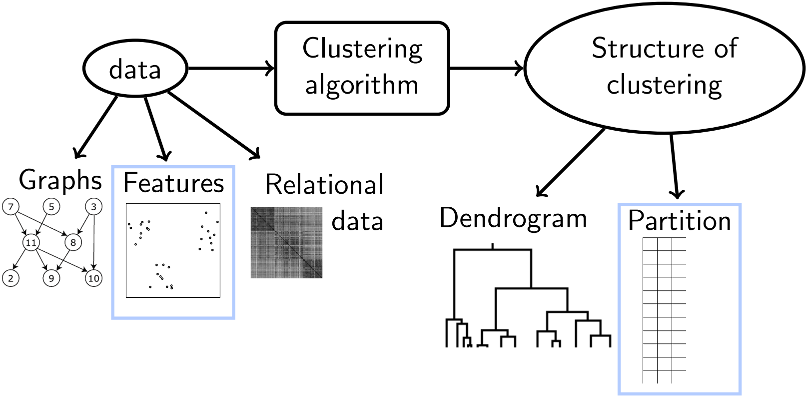
Clustering

Determine the group of objects based on a similarity notion



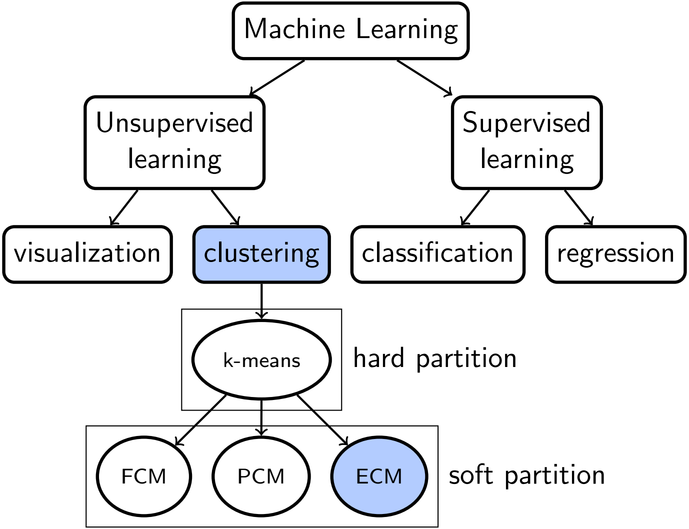
Clustering

Determine the group of objects based on a similarity notion



A famous clustering algorithm: k-means

Clustering : a technique of Machine Learning



Constrained clustering

Clustering problematic

No background knowledge

- how to define a similarity notion ?
- how to chose between several clustering solutions ?

Constrained clustering

Clustering problematic

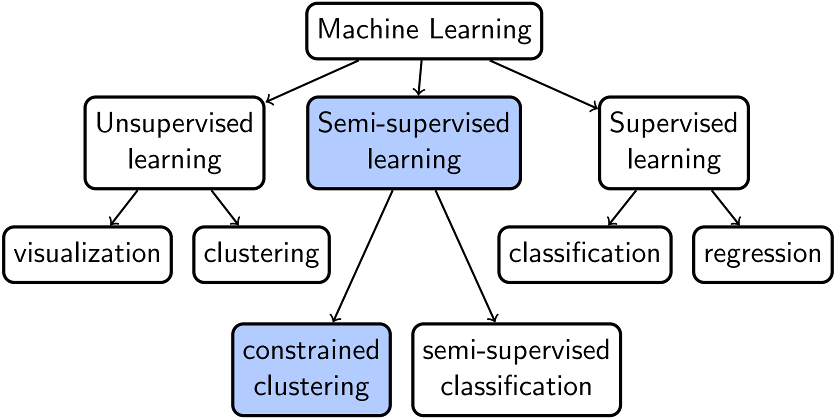
No background knowledge

- how to define a similarity notion ?
- how to chose between several clustering solutions ?

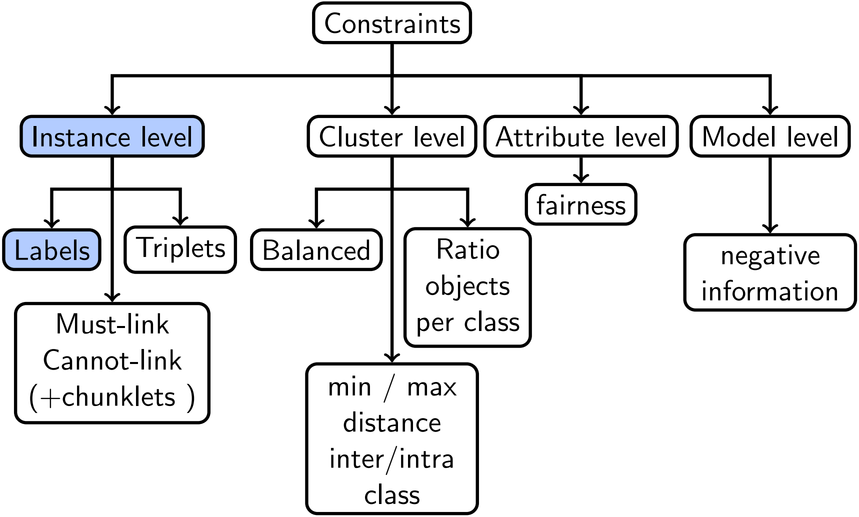
Expert information

- retrieve constraints from background knowledge
- semi-automatically collect constraints (active learning)

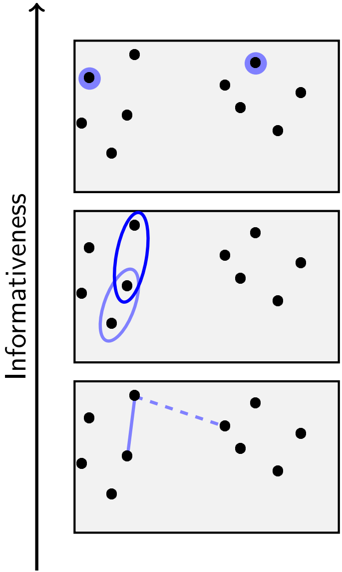
Constrained clustering



Constraint types



Instance level constraints



Label

An object belong to a class: $\mathbf{x}_i \in \mathcal{L}$

Triplet

$\mathbf{x}_a$ is closer to $\mathbf{x}_b$ than to $\mathbf{x}_c$: $d(x_a, x_b) < d(x_a, x_c)$

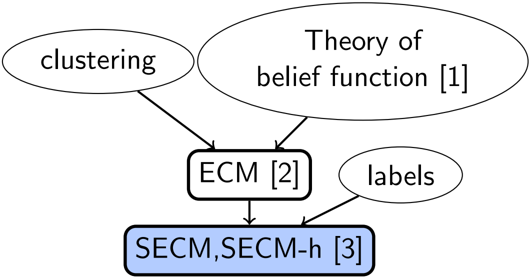
Must-Link / Cannot-Link:

Two objects are in the same/different class:

$(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{M}$ / $(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{C}$

Chunklets: set of objects in the same class

Motivations



[1] P. Smets, *The transferable belief model for quantified belief representation*, 1998

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

[3] V. Antoine & al, *Fast semi-supervised evidential clustering*, 2021

Outline : the soft variants of k-means

1. Background
   - FCM
   - ECM
2. SECM
   - Consistency measure
   - Objective function
   - Optimization
3. Experiments
4. Conclusion

Outline

1. Background
   - FCM
   - ECM
2. SECM
   - Consistency measure
   - Objective function
   - Optimization
3. Experiments
4. Conclusion

Fuzzy partition

- Each object has a degree of membership to each cluster
- $\mathbf{U} = (u_{ik})$ s.t $u_{ik} \in [0, 1], \sum_{k=1}^{c} u_{ik} = 1$

Example

Let ω_1 be the class of square, ω_2 the class of round

	p_{i1}	p_{i2}
	0	1
	1	0
	0.9	0.1
	0.5	0.5

Fuzzy c-means (FCM)

Geometrical model

- Each cluster ω_k is represented by a center $\mathbf{v}_k$
- Mahalanobis distance $d_{ik}^2 = (\mathbf{x}_i - \mathbf{v}_k)\mathbf{S}_k(\mathbf{x}_i - \mathbf{v}_k)$

Objective function

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

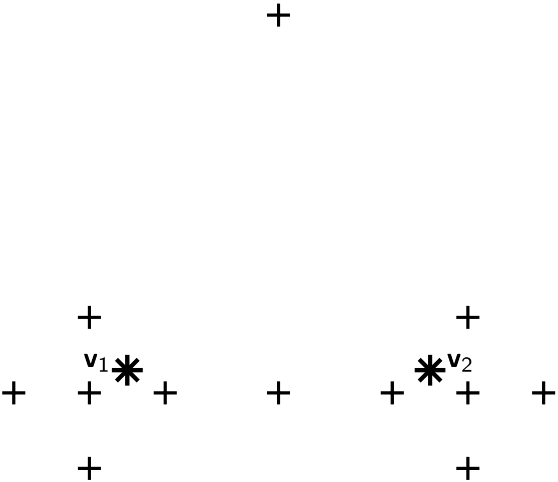
Subject to

$$\sum_{k=1}^{C} u_{ik} = 1 \text{ and } u_{ik} \geq 0 \quad \forall i, k$$

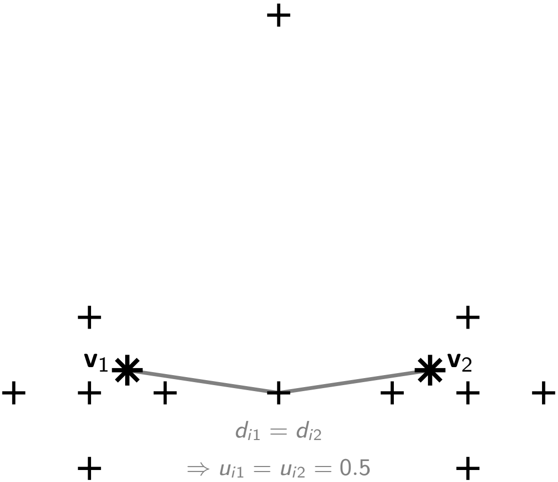
Gauss-Seidel optimization method

$$\min_{\mathbf{U}} J_{FCM} \quad \rightarrow \quad \min_{\mathbf{V}} J_{FCM} \quad \rightarrow \quad \min_{\mathbf{S}} J_{FCM} \quad \rightarrow \ldots$$

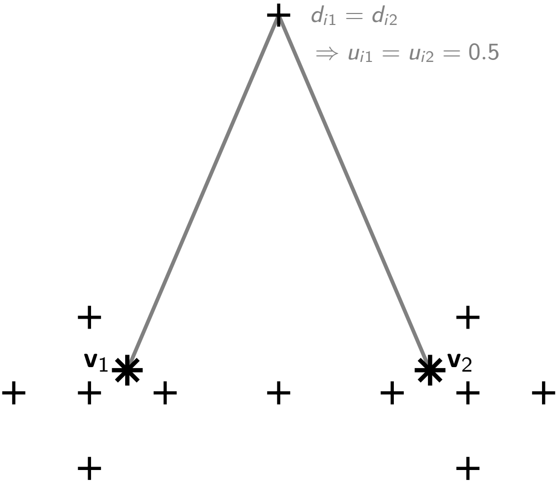
Problematic: imprecise assignments and outliers



Problematic: imprecise assignments and outliers



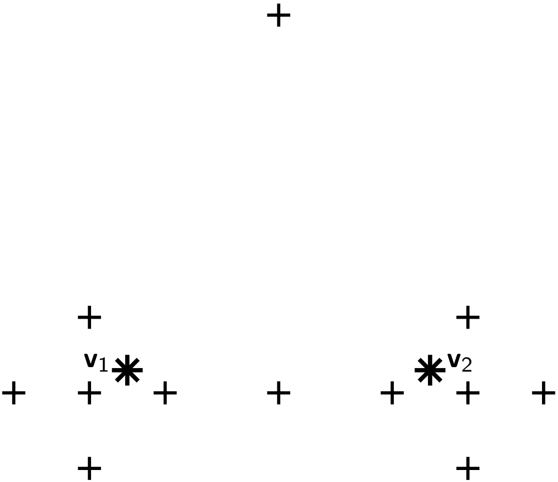
Problematic: imprecise assignments and outliers



Problematic: imprecise assignments and outliers

$u_{i1} = 0.3, \; u_{i2} = 0.7$ +

closed world ?



Belief function theory

Let Y be a variable taking values in a finite set Ω.

Mass function $m : 2^{\Omega} \to [0,1]$

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

- $m(\mathcal{A})$: degree of belief specific to $Y \in \mathcal{A}$
- If $m(\mathcal{A}) > 0$ then $\mathcal{A}$ is a focal set

Credal partition

- Each object has a degree of belief to each subset $\mathcal{A}_j \subseteq \Omega$
- $\mathbf{M} = (m_{ij})$ s.t $m_{ij} \in [0,1]$, $\sum_{\mathcal{A}_j \subseteq \Omega} m_{ij} = 1$

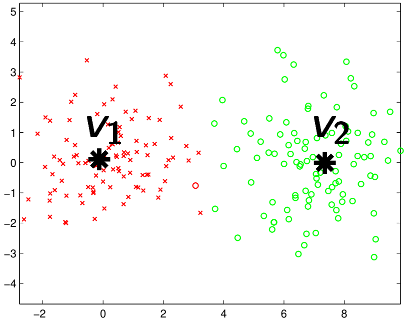
Example

Let ω_1 be the class of square, ω_2 the class of round

	$m_{i\emptyset}$	$m_{i\omega_1}$	$m_{i\omega_2}$	$m_{i\Omega}$
	0	0	1	0
	0	1	0	0
	0	0.9	0.1	0
	0	0	0	1
	1	0	0	0

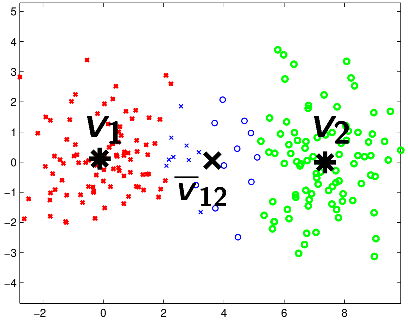
Evidential c-means (ECM)

- Each cluster ω_k is represented by a center $\mathbf{v}_k$
- Centroid $\bar{\mathbf{v}}_j$: barycenter of centers associated to classes composing $\mathcal{A}_j \subseteq \Omega$
- Distance d_{ij}^2 between $\mathbf{x}_i$ and $\bar{\mathbf{v}}_j$



Evidential c-means (ECM)

- Each cluster ω_k is represented by a center $\mathbf{v}_k$
- Centroid $\overline{\mathbf{v}}_j$: barycenter of centers associated to classes composing $\mathcal{A}_j \subseteq \Omega$
- Distance d_{ij}^2 between $\mathbf{x}_i$ and $\overline{\mathbf{v}}_j$



Evidential c-means (ECM)

Objective function

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

Subject to

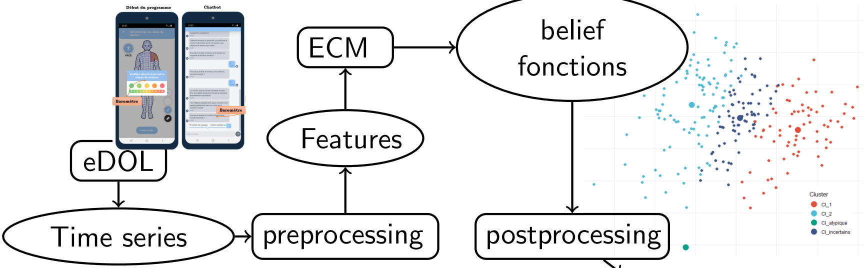
$$\sum_{\mathcal{A}_j \subseteq \Omega,\ \mathcal{A}_j \neq \emptyset} m_i(\mathcal{A}_j) + m_i(\emptyset) = 1,\ m_i(\mathcal{A}_j) \geq 0 \quad \forall i, j,$$

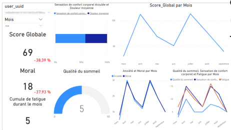
$$\det(\mathbf{S}_k) = 1 \quad \forall \omega_k \in \Omega$$

Gauss-Seidel optimization method

$$\text{opt}(\mathbf{M}) \rightarrow \text{opt}(\mathbf{V}) \rightarrow \text{opt}(\mathcal{S}) \rightarrow \ldots$$

Interest of ECM: application for health care [1]





compute statistical attributes (mean, std,...), selection of the most important variables

Facteurs associés	Cluster 1, $N = 71^1$	Cluster 2, $N = 102^1$	$p - value^2$	Total, N = 173
Age			0.3	
18 et 30 ans exclus	1 (1.4%)	5 (4.9%)		6 (3.5%)
30 et 50 ans inclus	45 (63%)	55 (54%)		100 (58%)
Plus 50 ans	25 (35%)	42 (41%)		67 (39%)
Nombre enfants			0.8	
1 et 2 enfants	32 (45%)	45 (44%)		77 (45%)
Pas d'enfants	29 (41%)	46 (45%)		75 (43%)
Plus 2 enfants	10 (14%)	11 (11%)		21 (12%)
Niveau Etudes			0.8	
BAC+4 ou +5	4 (5.6%)	7 (6.9%)		11 (6.4%)
BAC	12 (17%)	21 (21%)		33 (19%)
Intermédiaire	55 (77%)	74 (73%)		129 (75%)
Fumeur			0.2	
Non	51 (72%)	82 (80%)		133 (77%)
Oui	20 (28%)	20 (20%)		40 (23%)
Alcoolémie			0.9	
0 verre	18 (25%)	30 (29%)		48 (28%)
0 et 10 verres	39 (55%)	52 (51%)		91 (53%)
11 et 35 verre	12 (17%)	18 (18%)		30 (17%)
Plus 35	2 (2.8%)	2 (2.0%)		4 (2.3%)
kinésophobie			0.6	
Non	28 (39%)	36 (35%)		64 (37%)
Oui	43 (61%)	66 (65%)		109 (63%)
Fonctions Cognitives			0.065	
Anormales	50 (70%)	84 (82%)		134 (77%)
Normales	21 (30%)	18 (18%)		39 (23%)

$n(\%)^1$, Test exact de Fisher ; Test du Chi-carré de Pearson^2

[1] Armel Soubeiga & al, *Classification automatique de séries chronologiques de patients souffrant de douleurs chroniques*, EGC 2023

Outline

Background knowledge

Expert provides imprecise labels $\mathcal{A}_j$: $(\mathbf{x}_i, \mathcal{A}_j) \in \mathcal{L}$

Example of expert annotation

ω_1 for square, ω_2 for round, ω_3 for pentagon

	ω_1	ω_2	ω_3	$\mathcal{A}_j$
○	✓	✗	✗	ω_1
□	✗	✓	✗	ω_2
⬠	?	?	✗	$\omega_{12} = \{\omega_1, \omega_2\}$

Consistency between labels and hard credal partition

	credal partition							label	
	$m_{i\omega_1}$	$m_{i\omega_2}$	$m_{i\omega_{12}}$	$m_{i\omega_3}$	$m_{i\omega_{13}}$	$m_{i\omega_{23}}$	Ω	$\mathcal{A}_j$	
○	1	0	0	0	0	0	0	ω_1	++
○	0	0	1	0	0	0	0	ω_1	+
○	0	0	0	0	0	0	1	ω_1	=
○	0	1	0	0	0	0	0	ω_1	-

Consistency between labels and hard credal partition

	credal partition							label	
	$m_{i\omega_1}$	$m_{i\omega_2}$	$m_{i\omega_{12}}$	$m_{i\omega_3}$	$m_{i\omega_{13}}$	$m_{i\omega_{23}}$	Ω	$\mathcal{A}_j$	
○	1	0	0	0	0	0	0	ω_1	++
○	0	0	1	0	0	0	0	ω_1	+
○	0	0	0	0	0	0	1	ω_1	=
○	0	1	0	0	0	0	0	ω_1	-
◗	0	1	0	0	0	0	0	ω_{12}	++
◗	0	0	1	0	0	0	0	ω_{12}	+
◗	0	0	0	0	1	0	0	ω_{12}	=
◗	0	0	0	0	0	0	1	ω_{12}	=
◗	0	0	0	1	0	0	0	ω_{12}	-

Consistency between labels and hard credal partition

	credal partition							label		
	$m_{i\omega_1}$	$m_{i\omega_2}$	$m_{i\omega_{12}}$	$m_{i\omega_3}$	$m_{i\omega_{13}}$	$m_{i\omega_{23}}$	Ω	$\mathcal{A}_j$		T_{ij}
	1	0	0	0	0	0	0	ω_1	++	1
	0	0	1	0	0	0	0	ω_1	+	1/2
	0	0	0	0	0	0	1	ω_1	=	1/3
	0	1	0	0	0	0	0	ω_1	-	0
	0	1	0	0	0	0	0	ω_{12}	++	1
	0	0	1	0	0	0	0	ω_{12}	+	$\sqrt{2}/2$
	0	0	0	0	1	0	0	ω_{12}	=	1/2
	0	0	0	0	0	0	1	ω_{12}	=	$\sqrt{2}/3$
	0	0	0	1	0	0	0	ω_{12}	-	0

Consistency measure

$$T_{ij} = T_i(\mathcal{A}_j) = \sum_{\mathcal{A}_\ell \cap \mathcal{A}_j \neq \emptyset} \frac{|\mathcal{A}_j \cap \mathcal{A}_\ell|^{r/2}}{|\mathcal{A}_\ell|^r} m_{i\ell}, \; r \geq 0 \text{ a constant}$$

Seed evidential clustering: SECM

	credal partition							label		
	$m_{i\omega_1}$	$m_{i\omega_2}$	$m_{i\omega_{12}}$	$m_{i\omega_3}$	$m_{i\omega_{13}}$	$m_{i\omega_{23}}$	Ω	$\mathcal{A}_j$		T_{ij}
	1	0	0	0	0	0	0	ω_1	++	1
	0	1	0	0	0	0	0	ω_1	-	0
	0	1	0	0	0	0	0	ω_{12}	++	1
	0	0	0	1	0	0	0	ω_{12}	-	0

Basic idea

If $(\mathbf{x}_i, \mathcal{A}_j) \in \mathcal{L} \Rightarrow T_{ij}$ should be high

Objective function to minimize

$$J_{SECM} = J_{ECM} + \gamma \sum_{(\mathbf{x}_i, \mathcal{A}_j) \in \mathcal{L}} 1 - T_{ij}$$

Consistency measure: r study

Consistency measure

$$T_{ij} = \sum_{\mathcal{A}_\ell \cap \mathcal{A}_j \neq \emptyset} \frac{|\mathcal{A}_j \cap \mathcal{A}_\ell|^{r/2}}{|\mathcal{A}_\ell|^r} m_{i\ell}$$

$r = 1$

$$T_{ij} = \sum_{\mathcal{A}_\ell \cap \mathcal{A}_j \neq \emptyset} \frac{|\mathcal{A}_j \cap \mathcal{A}_\ell|^{1/2}}{|\mathcal{A}_\ell|} m_{i\ell}$$

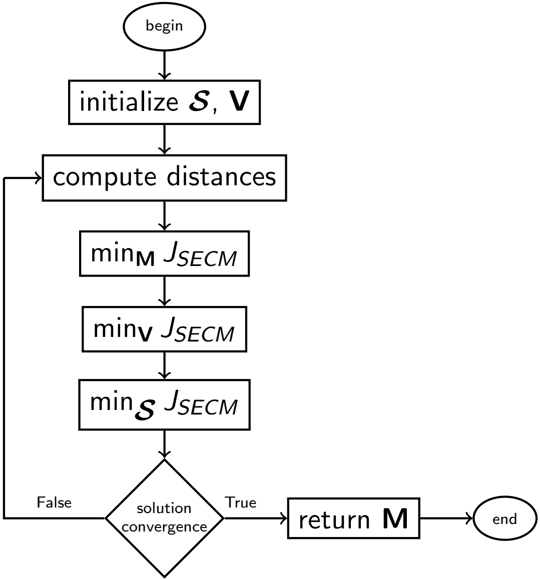
- Low cardinality are favored
- \+ make decision when labels are known

$r = 0$

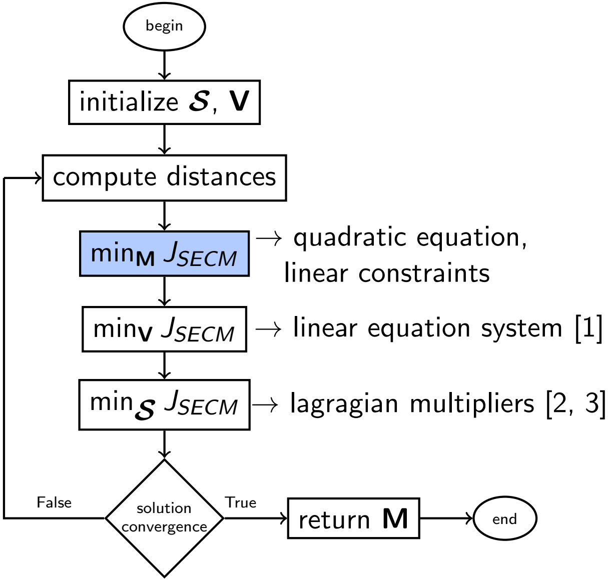
$$T_{ij} = \sum_{\mathcal{A}_\ell \cap \mathcal{A}_j \neq \emptyset} m_{i\ell} = pl_i(\mathcal{A}_j)$$

- no difference between low and high cardinality
- \+ robust to noisy labels

Gauss-Seidel optimization method



Gauss-Seidel optimization method

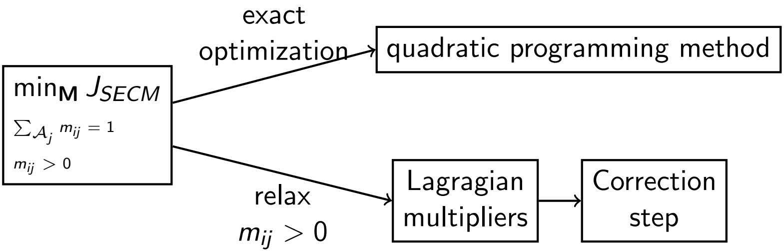


[1] M.-H. Masson & al, *ECM: An evidential version of the fuzzy c-means algorithm*, 2008

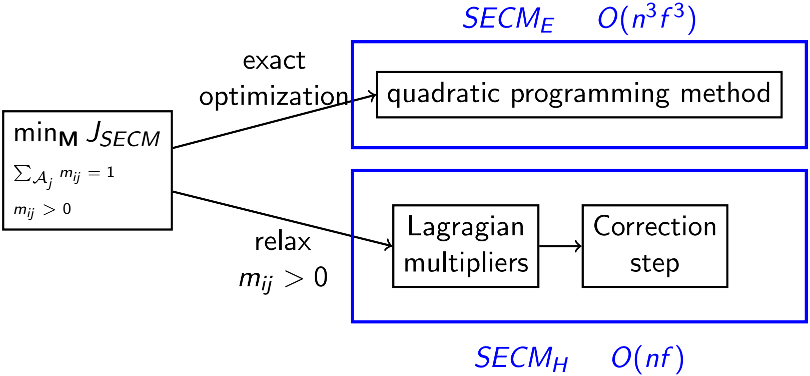
[2] D. Gustafson & al, *Fuzzy clustering with a fuzzy covariance matrix*, 1978

[3] V. Antoine & al, *CECM:constrained evidential c-means algorithm*, 2012

Optimization of the credal partition



Optimization of the credal partition



Heuristic optimization $SECM_H$

Hypothesis

Relaxing $m_{ij} > 0$ has an insignificant impact on the solution

Lagragian multipliers

$$\mathcal{L}(\mathbf{M}, \lambda_1, \ldots, \lambda_n) = J_{ECM} + \left(\gamma \sum_{(\mathbf{x}_i, \mathcal{A}_j) \in \mathcal{L}} 1 - T_{ij} \right) - \sum_{i=1}^{n} \lambda_i \left(\sum_{\mathcal{A}_j} m_{ij} - 1 \right)$$

$$\frac{\partial \mathcal{L}}{\partial m_{ij}} = 0 \Rightarrow m_{ij} = \frac{1}{|\mathcal{A}_j|^\alpha d_{ij}^2 D} + \gamma f(\mathbf{x}_i, \mathcal{A}_j) - \gamma g(\mathbf{x}_i, \mathcal{A}_\ell)$$

Heuristic optimization $SECM_H$

Hypothesis

Relaxing $m_{ij} > 0$ has an insignificant impact on the solution

Lagragian multipliers

$$\mathcal{L}(\mathbf{M}, \lambda_1, \ldots, \lambda_n) = J_{ECM} + \left(\gamma \sum_{(\mathbf{x}_i, \mathcal{A}_j) \in \mathcal{L}} 1 - T_{ij} \right) - \sum_{i=1}^{n} \lambda_i \left(\sum_{\mathcal{A}_j} m_{ij} - 1 \right)$$

if $(\mathbf{x}_i, \mathcal{A}_j) \in \mathcal{L}$

$$\frac{\partial \mathcal{L}}{\partial m_{ij}} = 0 \Rightarrow m_{ij} = \boxed{\frac{1}{|\mathcal{A}_j|^\alpha d_{ij}^2 D}} + \boxed{\gamma f(\mathbf{x}_i, \mathcal{A}_j)} - \boxed{\gamma g(\mathbf{x}_i, \mathcal{A}_\ell)}$$

ECM update formula

if $(\mathbf{x}_i, \mathcal{A}_\ell) \in \mathcal{L}$

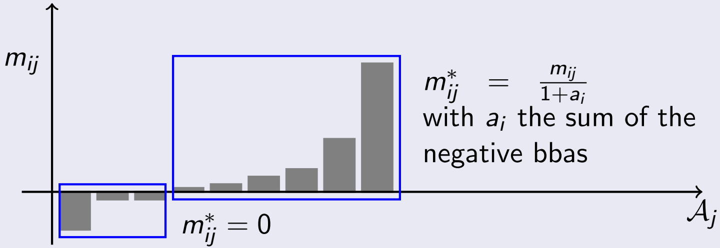
Heuristic optimization $SECM_H$

$$m_{ij} = \underbrace{\frac{1}{|\mathcal{A}_j|^{\alpha} d_{ij}^2 D}}_{\geq 0} + \underbrace{\gamma f(\mathbf{x}_i, \mathcal{A}_j)}_{\geq 0} - \underbrace{\gamma g(\mathbf{x}_i, \mathcal{A}_\ell)}_{\geq 0}$$

Hence,

- $m_{ij} \in]-\infty, 1]$
- $\sum_{\mathcal{A}_j \subseteq \Omega} m_{ij} = 1$

Correction step for $\mathbf{x}_i$



$m_{ij}^* = \frac{m_{ij}}{1+a_i}$
with a_i the sum of the negative bbas

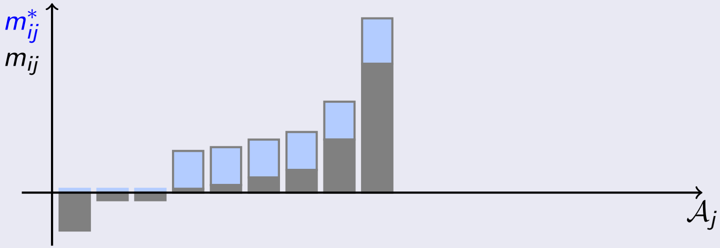
Heuristic optimization $SECM_H$

$$m_{ij} = \underbrace{\boxed{\frac{1}{|\mathcal{A}_j|^\alpha d_{ij}^2 D}}}_{\geq 0} + \underbrace{\boxed{\gamma f(\mathbf{x}_i, \mathcal{A}_j)}}_{\geq 0} - \underbrace{\boxed{\gamma g(\mathbf{x}_i, \mathcal{A}_\ell)}}_{\geq 0}$$

Hence,

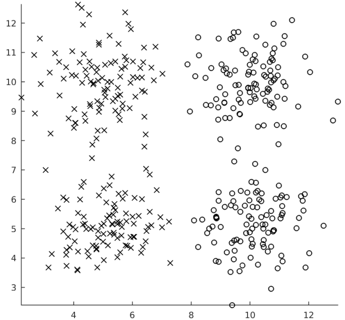
- $m_{ij} \in]-\infty, 1]$
- $\sum_{\mathcal{A}_j \subseteq \Omega} m_{ij} = 1$

Correction step for $\mathbf{x}_i$

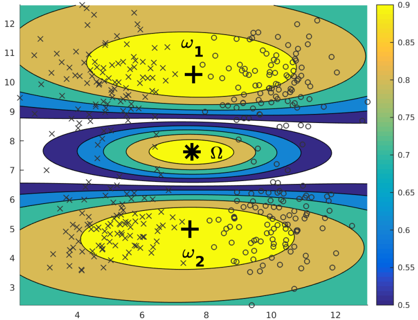


Outline

Constraints interest

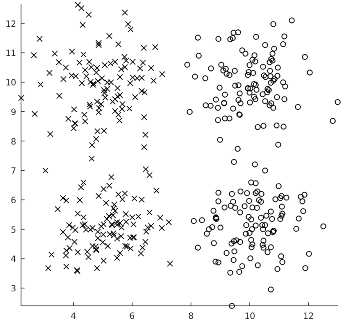


gaussK2

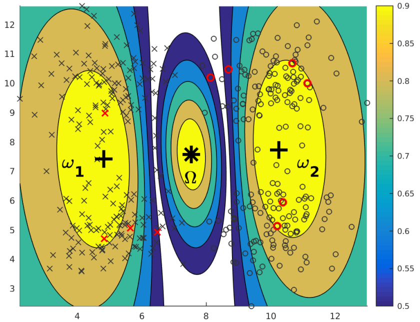


ECM

Constraints interest



gaussK2



SECM

Experimental protocol

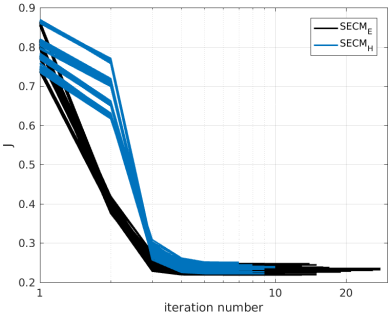
Data sets

	# objects	# attributes	# classes
Column	310	6	3
Iris	150	4	3
Wine	178	13	3

Evaluation method based on true known classes

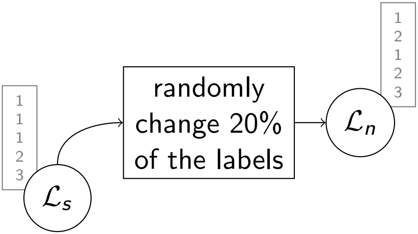
- random constraints selection
- evaluation measure:
  - pignistic transformation $\Rightarrow$ fuzzy partition
  - maximum of probability $\Rightarrow$ hard partition
  - ARI $\in [0, 1]$

Optimization analysis on Wine data set



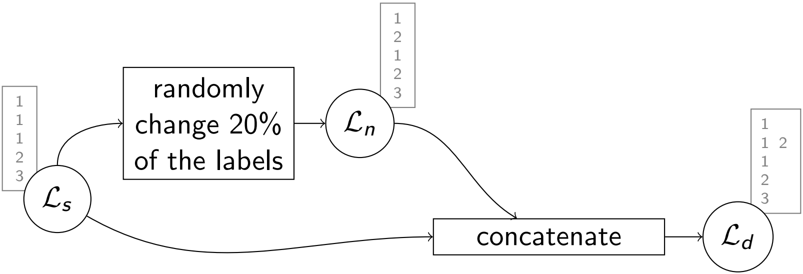
30 const.	$SECM_H$	$SECM_E$
J_{SECM} $(\times 10^{-3})$	236.3[1.1]	232.7[1.1]
CPU (s)	0.19[0.00]	0.89[0.03]
ARI	0.92[0.02]	0.92[0.03]

Influence of the r parameter on Iris data set



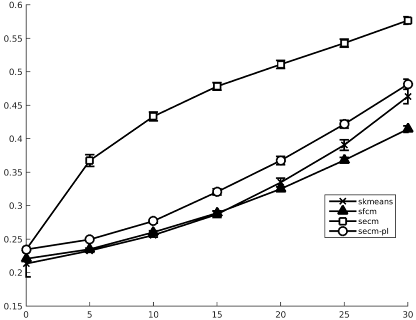
ARI [std]	r	$\mathcal{L}_s$ set		$\mathcal{L}_n$ set	
		SECM$_H$	*SECMpl$_H$*	*SECM$_H$*	*SECMpl$_H$*
	0	0.67 [0.01]	0.67 [0.01]	0.67 [0.01]	0.67 [0.01]
	10	**0.82** [0.07]	0.77 [0.07]	0.51 [0.14]	**0.62** [0.08]
	20	**0.90** [0.05]	0.86 [0.06]	0.58 [0.10]	**0.61** [0.08]
	30	**0.92** [0.03]	0.89 [0.05]	0.59 [0.08]	0.58 [0.07]

Influence of the r parameter on Iris data set

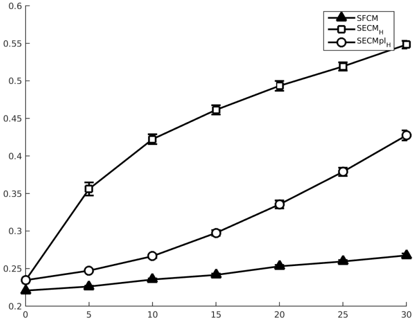


		$\mathcal{L}_s$ set		$\mathcal{L}_n$ set	
		$SECM_H$	$SECMpl_H$	$SECM_H$	$SECMpl_H$
ARI [std]	0	0.67 [0.01]	0.67 [0.01]	0.67 [0.01]	0.67 [0.01]
	10	**0.82** [0.07]	0.77 [0.07]	0.51 [0.14]	**0.62** [0.08]
	20	**0.90** [0.05]	0.86 [0.06]	0.58 [0.10]	**0.61** [0.08]
	30	**0.92** [0.03]	0.89 [0.05]	0.59 [0.08]	0.58 [0.07]

Algorithm comparison on Column data set



single labels $\mathcal{L}_s$



double labels $\mathcal{L}_d$

Outline

Conclusion

SECM

- evidential clustering
- incorporation of labels
- \+ credal partition is full of information
- \+ labels improve performances
- \- computational complexity
- \- sensitivity to label selection

Thank you