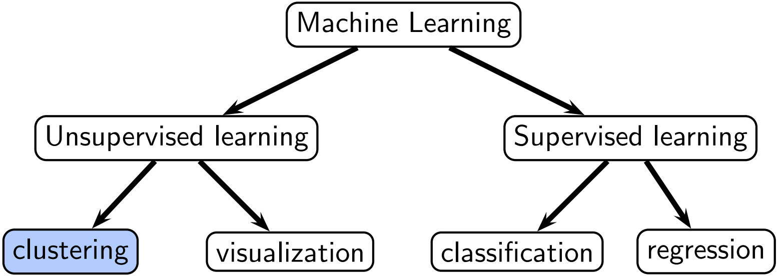
Soft clustering: a review of k-means variants

V. Antoine

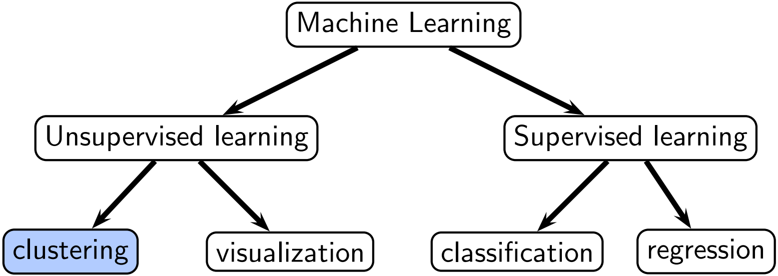
Clermont Auvergne University, LIMOS, UMR CNRS 6158, France

September 2018

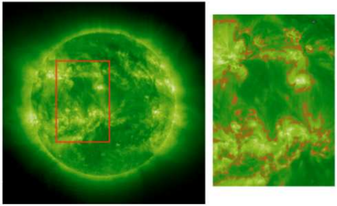
Clustering : a technique of Machine Learning



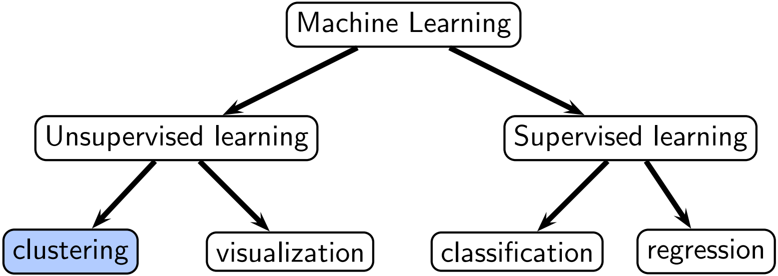
Clustering : a technique of Machine Learning



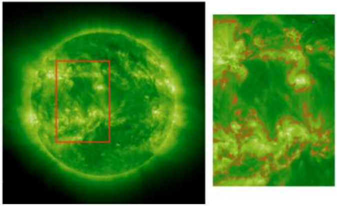
Astronomy [1]

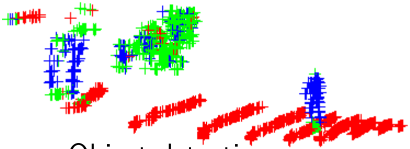


Clustering : a technique of Machine Learning



Astronomy [1]





Object detection

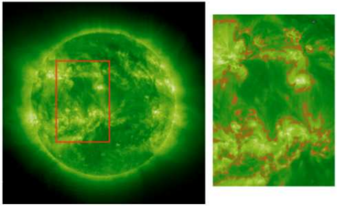
Clustering : a technique of Machine Learning

Machine

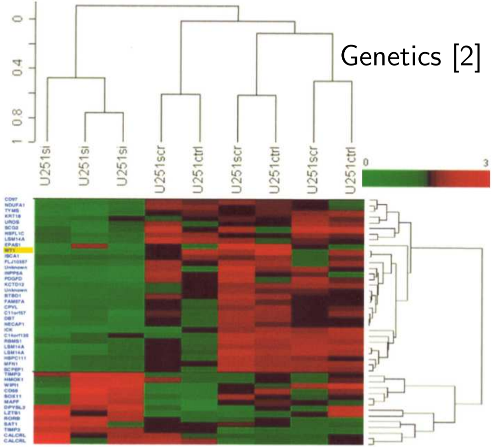
Unsupervised learning

clustering | visualization

Astronomy [1]



Genetics [2]

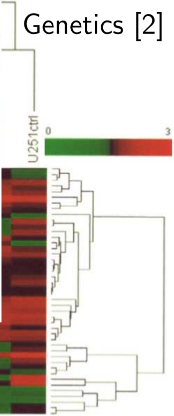


Object detection

Clustering : a technique of Machine Learning

Text mining

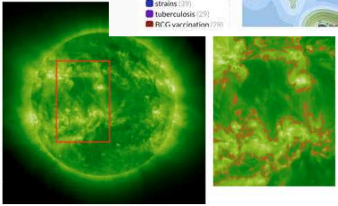
Genetics [2]



Unsuperv

clustering

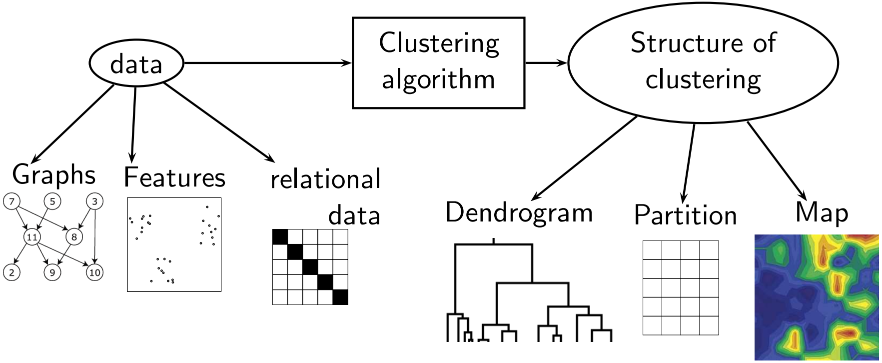
Astronomy [1]



Object detection

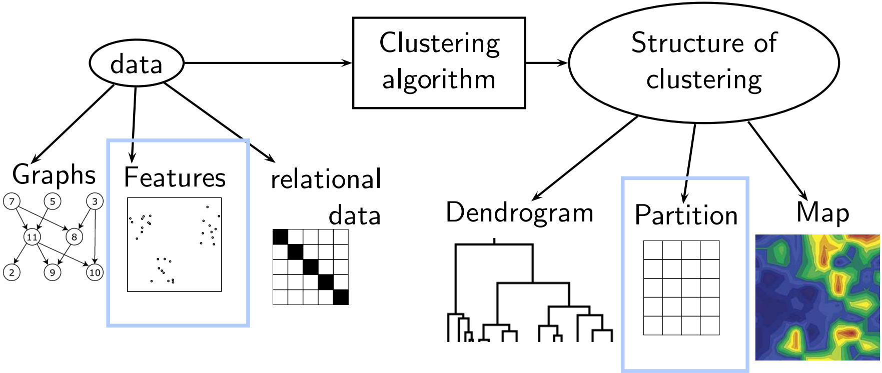
Clustering

Determine the group of objects following a similarity notion



Clustering

Determine the group of objects following a similarity notion



Partition types

- Let $\mathbf{X} = (\mathbf{x}_i)$ be a collection of objects s.t. $\mathbf{x}_i \in \mathbb{R}^p$,
- $\Omega = \{\omega_1 \ldots \omega_c\}$ a set of c clusters,

Hard and soft partitions:

- hard/crisp partition
- fuzzy partition
- possibilistic partition
- rough partition
- credal partition

Outline : the soft variants of k-means

Outline

Hard partition

- Each object is assigned to one and only one cluster
- $\mathbf{P} = (p_{ik})$ s.t $p_{ik} \in \{0, 1\}$, $\sum_{k=1}^{c} p_{ik} = 1$

Example

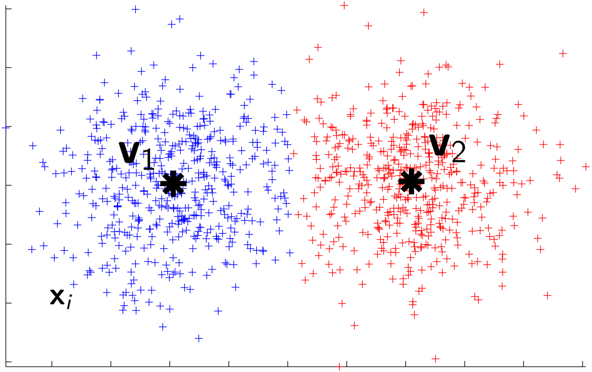
Let ω_1 be the class of square, ω_2 the class of round

	p_{i1}	p_{i2}
○	0	1
□	1	0
□	1	0

k-means

Geometrical model:

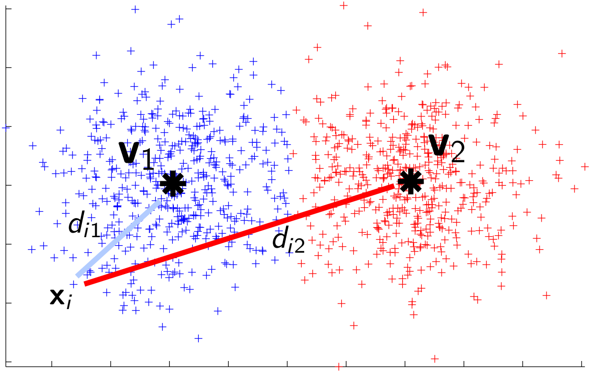
- Each cluster ω_k is represented by a center $\mathbf{v}_k$
- Euclidean distance $d_{ik}^2 = \|\mathbf{x}_i - \mathbf{v}_k\|^2$



k-means

Geometrical model:

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



k-means

Objective function

$$J_{KM} = \sum_{i=1}^{N} \sum_{k=1}^{c} p_{ik} d_{ik}^2$$

Subject to

$$\sum_{k=1}^{c} p_{ik} = 1 \text{ and } p_{ik} \in \{0,1\} \forall i,k$$

Optimization

NP-Hard $\Rightarrow$ minimization using an iterative procedure:

$$\text{fix } \mathbf{V}, \min_{\mathbf{P}} J_{KM} \quad \rightleftarrows \quad \text{fix } \mathbf{P}, \min_{\mathbf{V}} J_{KM}$$

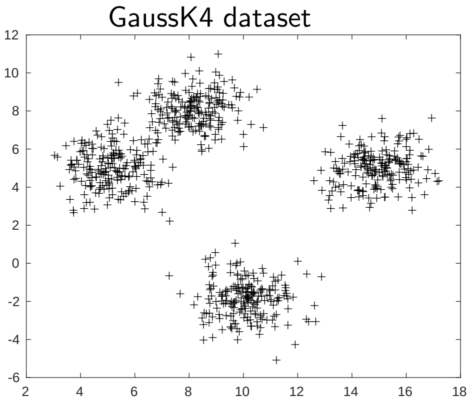
Advantage

Fast

Disadvantage

Risk of local minimum

Determination of the number of clusters

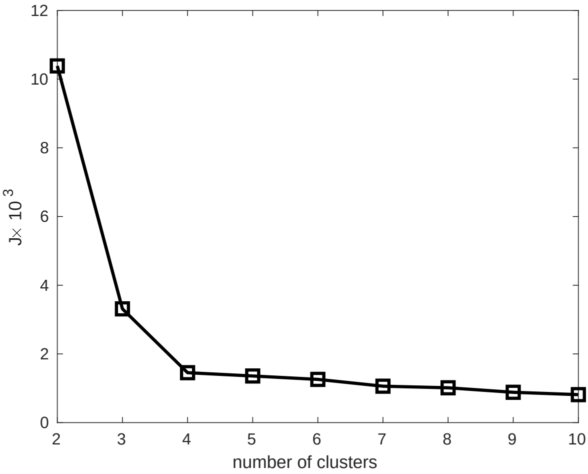


For c=1 to 10

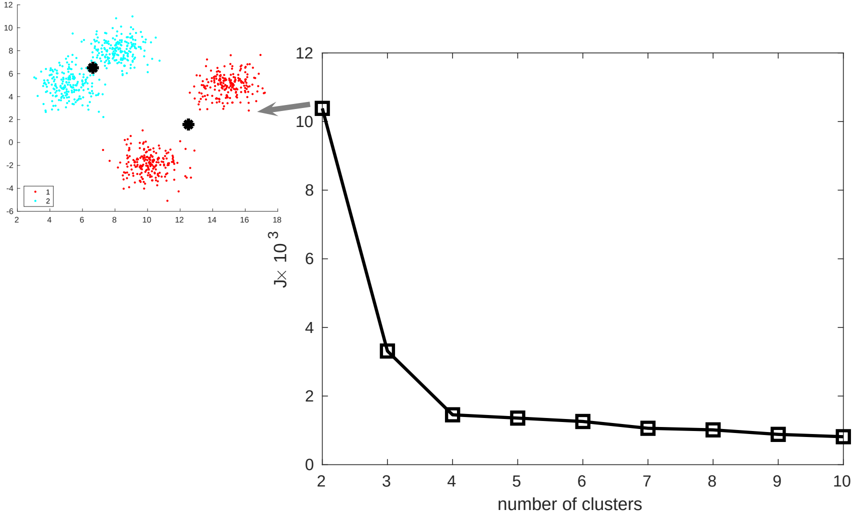
- run kmeans
- evaluate the partition

Plot evaluation measure vs number of clusters

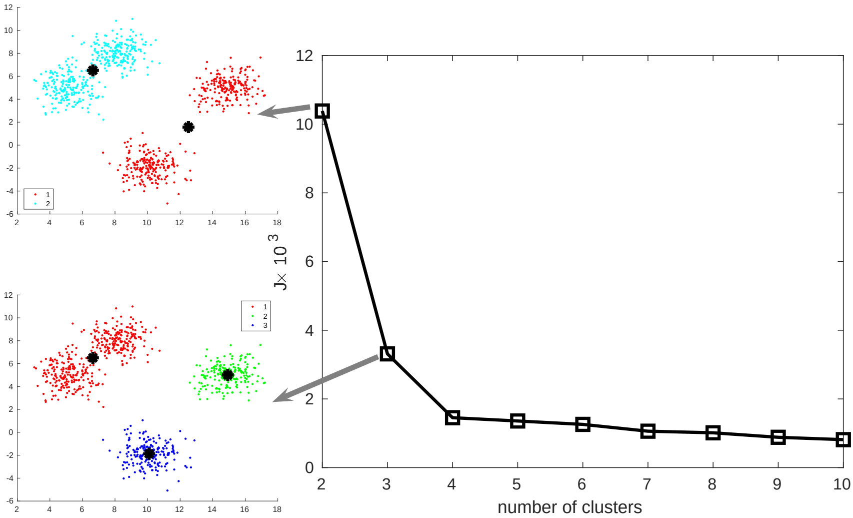
Determination of the number of clusters



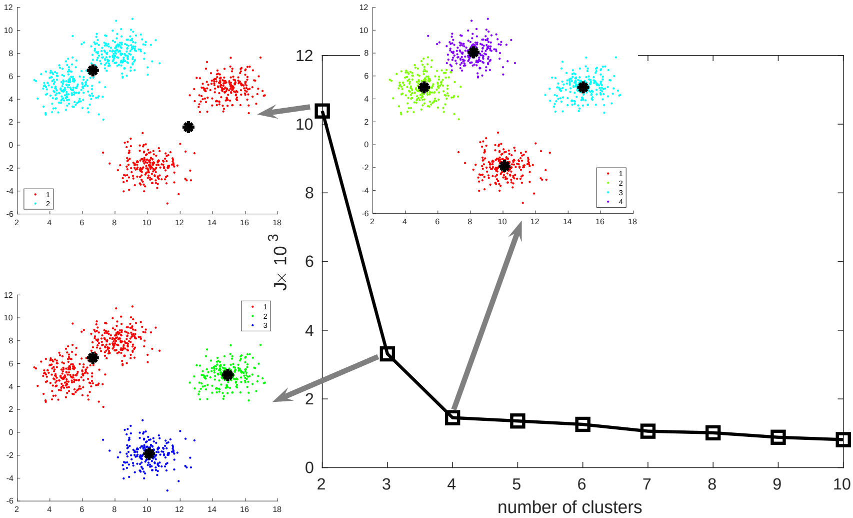
Determination of the number of clusters



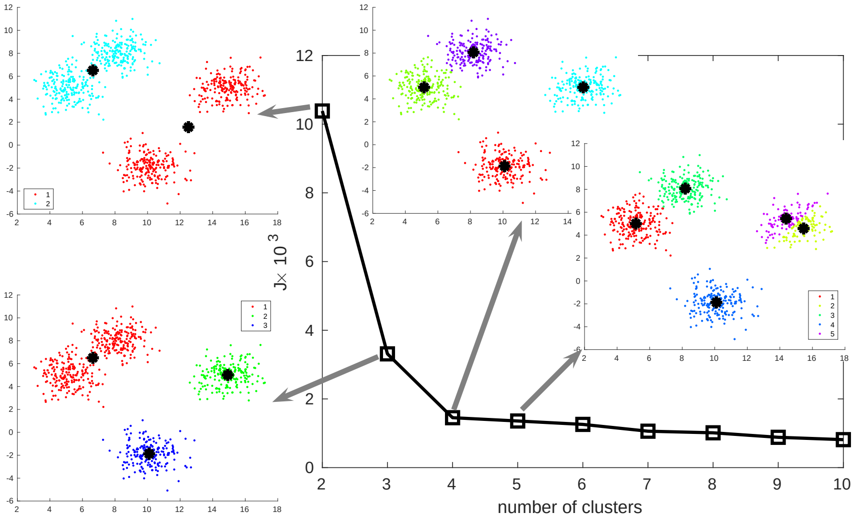
Determination of the number of clusters



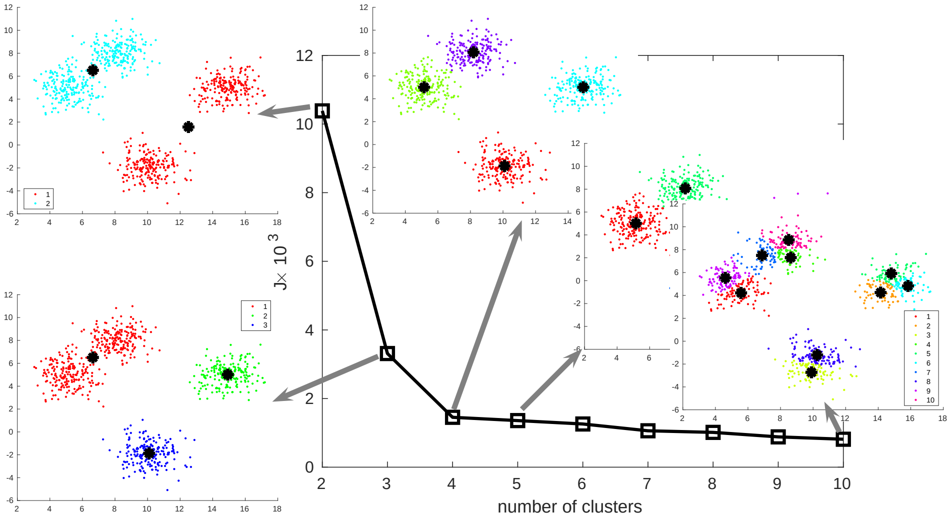
Determination of the number of clusters



Determination of the number of clusters

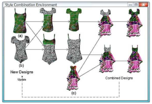


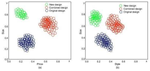
Determination of the number of clusters

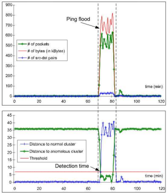


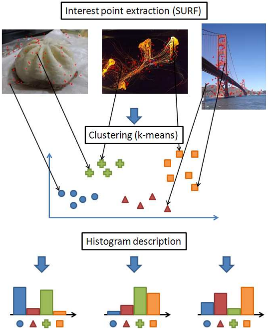
Applications

- Biology and Bioinformatics
  - Protein structure prediction
  - Gene expression
- Engineering
  - Encoding/decoding
  - Image retrieval system
  - Color image segmentation
- Business and Economics
  - Exploration of shopping orientations [10]
  - Marketing
- Faud detection [6]



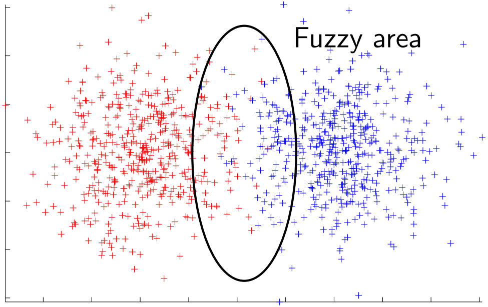






www.olivier-augereau.com/blog/?p=358

Problematic



Original data

Solution

Express uncertainty about the clustering result

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

Outline

Fuzzy c-means (FCM)

Objective function

$$J_{FCM} = \sum_{i=1}^{N} \sum_{k=1}^{C} u_{ik}^{\beta} d_{ik}^{2}$$

Subject to

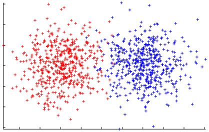
$$\sum_{k=1}^{C} u_{ik} = 1 \text{ and } u_{ik} \geq 0 \quad \forall i, k$$

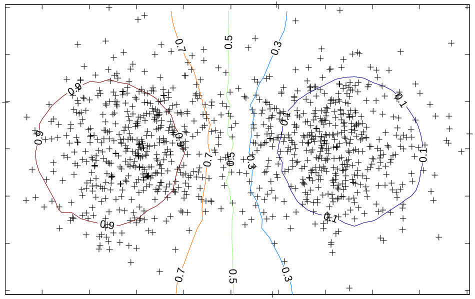
Aternate optimization

$$\min_{\mathbf{U}} J_{FCM} \quad \rightleftarrows \quad \min_{\mathbf{V}} J_{FCM}$$

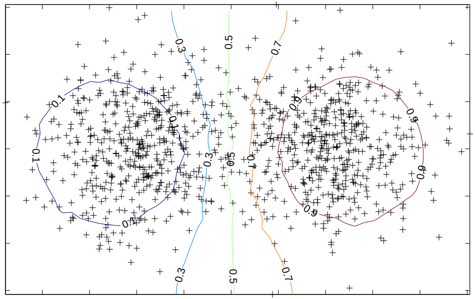
Fuzzy c-means (FCM)

Original data





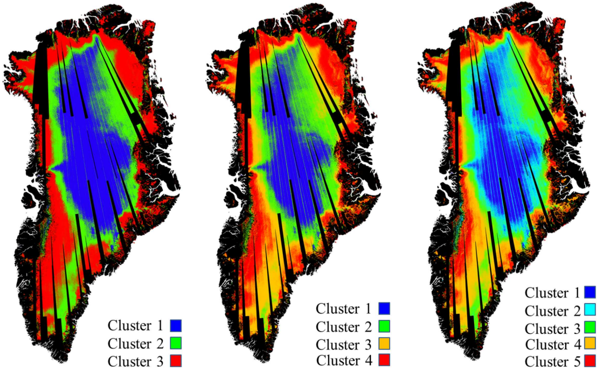
ω_1



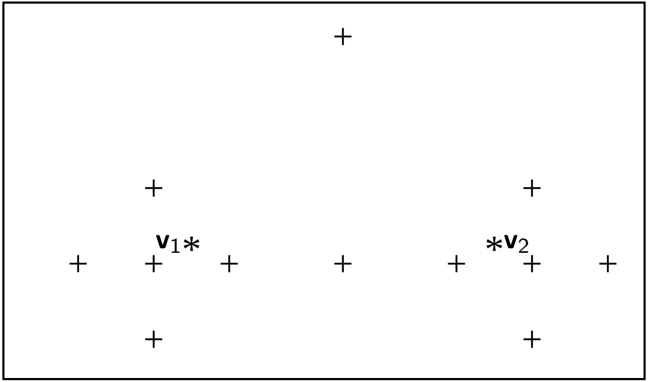
ω_2

Applications

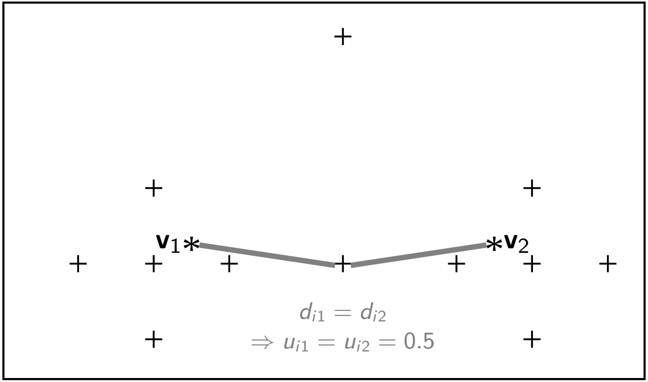
Same as k-means, but avoid decision threshold.
Example : Characterization of snow facies [8]



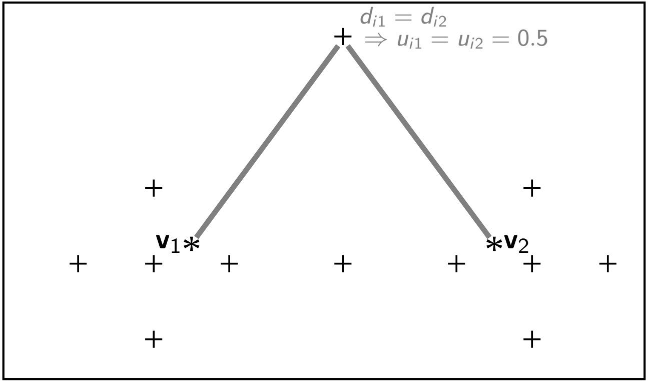
Problematic: outliers



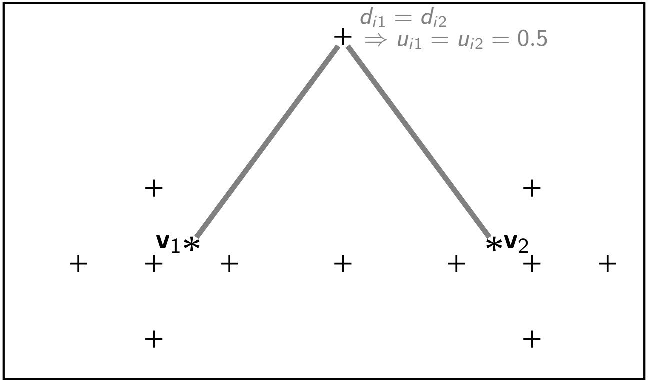
Problematic: outliers



Problematic: outliers



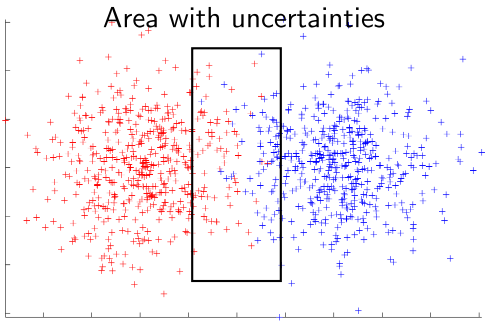
Problematic: outliers



Solution

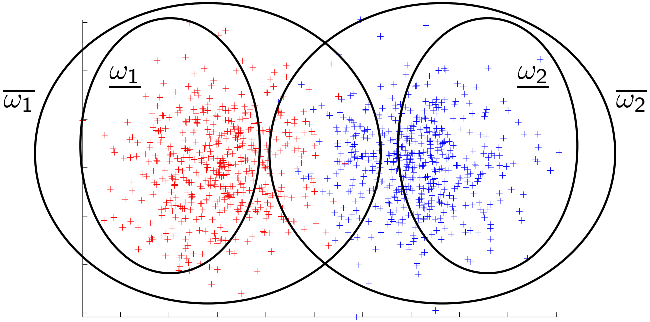
Relax the sum constraint

Problematic: where are the limits of an uncertain region ?



Original data

Problematic: where are the limits of an uncertain region ?



Original data

Solution

Express for each cluster ω_k a lower approximation $\underline{\omega}_k$ and an upper approximation $\overline{\omega}_k$

Rough partition

- Each object has a lower/upper approximation to each cluster
- $(\overline{\lambda}, \underline{\lambda}) = ((\overline{\lambda}_{ik}), (\underline{\lambda}_{ik}))$ s.t $\overline{\lambda}_{ik}, \underline{\lambda}_{ik} \in \{0, 1\}$,
  - if $\mathbf{x}_i \in \underline{\omega_k}$ then $\underline{\lambda}_{ik'} = 1$ and $\sum_{k=1}^{c} \overline{\lambda}_{ik} = 0$,
  - otherwise, $\sum_{k=1}^{c} \underline{\lambda}_{ik} = 0$ and $\sum_{k=1}^{c} \overline{\lambda}_{ik} \geq 1$.

Example

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

	$\underline{\lambda}_{i1}$	$\overline{\lambda}_{i1}$	$\underline{\lambda}_{i2}$	$\overline{\lambda}_{i2}$
	1	0	0	0
	0	0	1	0
	0	0	1	0
	0	1	0	1

Outline

Rough c-means

Objective function

$$J_{RCM} = \sum_{i=1}^{N} \sum_{k=1}^{C} \frac{\gamma}{\underline{n}_k} \underline{\lambda}_{ik} d_{ik}^2 + \frac{1-\gamma}{\overline{n}_k} \overline{\lambda}_{ik} d_{ik}^2$$

such that

- $\gamma \in [0\ 1]$ is a fixed weight
- $\underline{n}_k / \overline{n}_k$ are the number of objects in $\underline{\omega}_k / \overline{\omega}_k$

Subject to

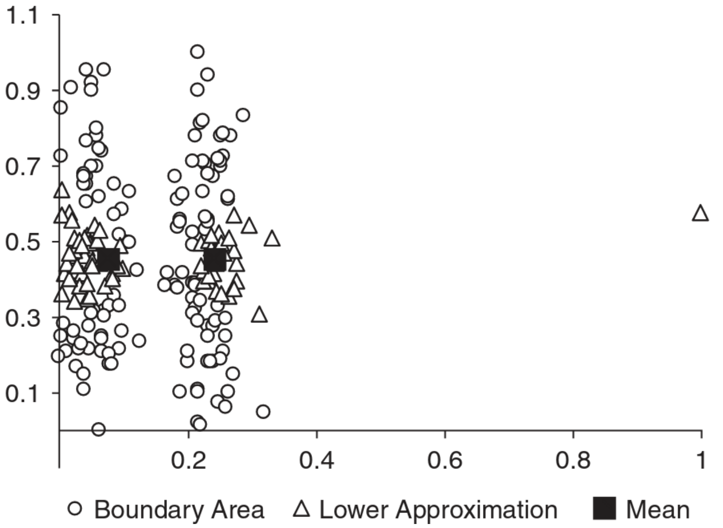
$$\left(\sum_{k=1}^{c} \underline{\lambda}_{ik} = 1 \vee \sum_{k=1}^{c} \underline{\lambda}_{ik} = 0 \right) \wedge \left(\sum_{k=1}^{c} \underline{\lambda}_{ik} = 0 \vee \sum_{k=1}^{c} \underline{\lambda}_{ik} \geq 2 \right) \text{ and}$$

$$\overline{\lambda}_{ik}, \underline{\lambda}_{ik} \in \{0, 1\} \quad \forall i, k$$

Aternate optimization

$$\min_{\underline{\lambda}, \overline{\lambda}} J_{RCM} \quad \rightleftarrows \quad \min_{\mathbf{V}} J_{FCM}$$

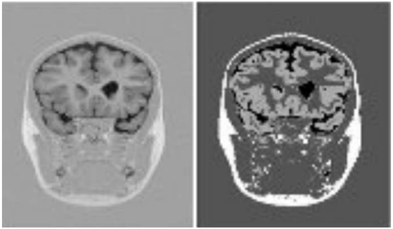
Rough c-means [7]

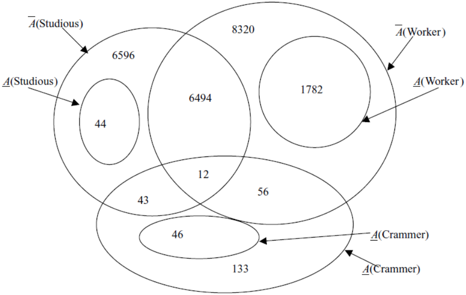


Applications

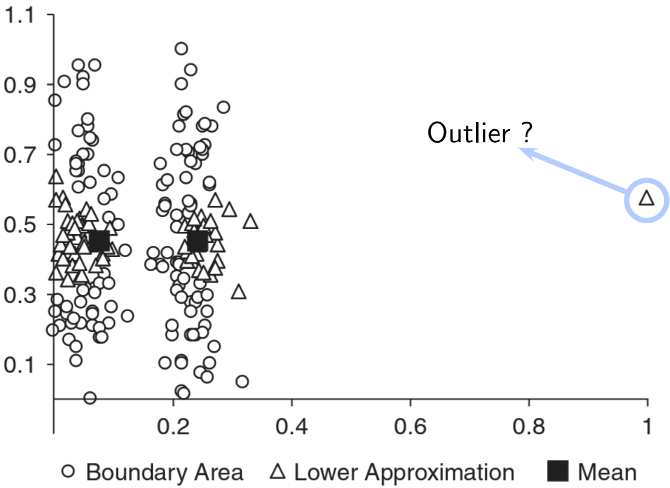
Great interest on boundary regions:

- Biology and Bioinformatics
- Medical imagery [5]
- Forest cover data
- Business and Economics
- Website profiles [4]
- . . .





Problematic



Possibilistic partition

- Each object has a degree of possibility to each cluster
- $\mathbf{T} = (t_{ik})$ s.t $t_{ik} \in [0, 1]$

Example

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

	t_{i1}	t_{i2}
	0	1
	1	0
	1	0.1
	1	1
	0	0

Possibilistic transformations

Possibilistic partition

	t_{i1}	t_{i2}
○	0	1
□	1	0
▢	1	0.1
D	1	1
☆	0	0

normalization

$$u_{ik} = \frac{t_{ik}}{\sum_{k=1}^{c} t_{ik}}$$

Fuzzy partition

	u_{i1}	u_{i2}
○	0	1
□	1	0
▢	0.91	0.09
D	0.5	0.5
☆	?	?

maximum

$$p_{ik} = \max_{k=1}^{c} t_{ik}$$

Hard partition

	p_{i1}	p_{i2}
○	0	1
□	1	0
▢	1	0
D	?	?
☆	?	?

Outline

Possibilistic c-means (PCM)

Objective function

$$J_{PCM} = \sum_{i=1}^{N} \sum_{k=1}^{C} t_{ik}^{\beta} d_{ik}^{2} + \sum_{k=1}^{c} \gamma_k \sum_{i=1}^{N} (1 - t_{ik})^m$$

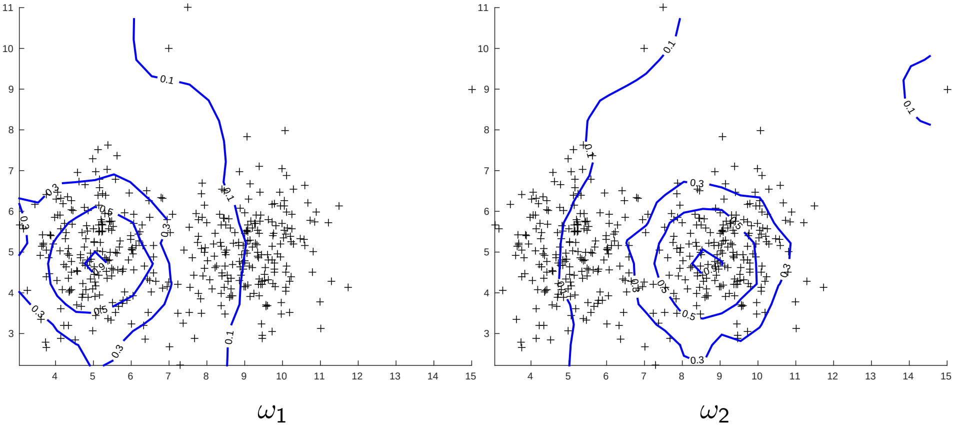
Subject to

$$t_{ik} \geq 0 \quad \forall i, k$$

Aternate optimization

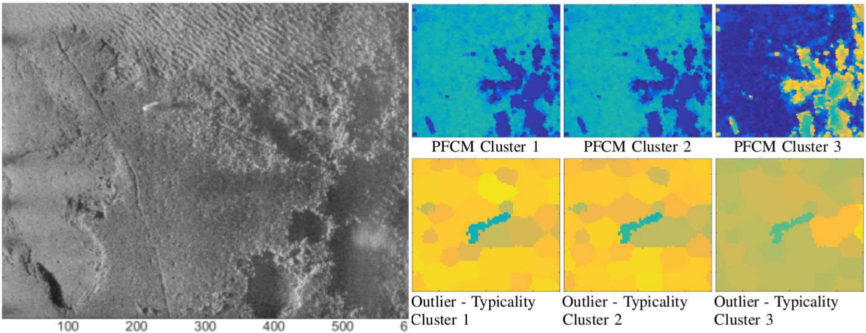
$$\min_{\mathbf{T}} J_{PCM} \quad \rightleftarrows \quad \min_{\mathbf{V}} J_{PCM}$$

Possibilistic c-means (PCM)



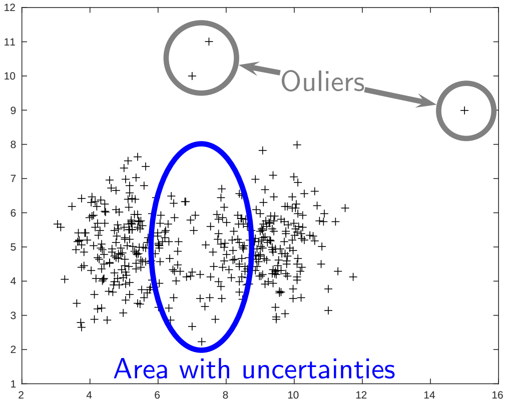
Applications

Same as FCM, but handles outliers.
Example : sonar image segmentation [11]



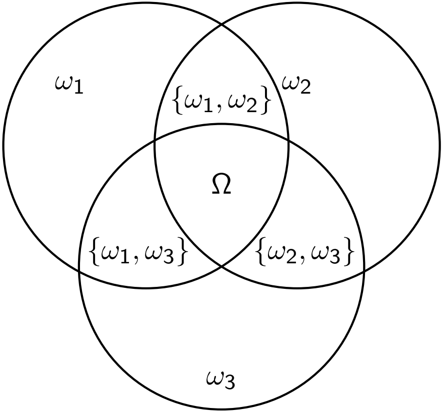
Problematic

Can we represent uncertain region and outliers with a mathematical framework ?



Credal partition

Uncertainties represented with subsets of $\Omega = \{\omega_1, \dots, \omega_c\}$



$\emptyset$

Credal partition

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

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

Derivative notions

Belief function

Total support given to A:

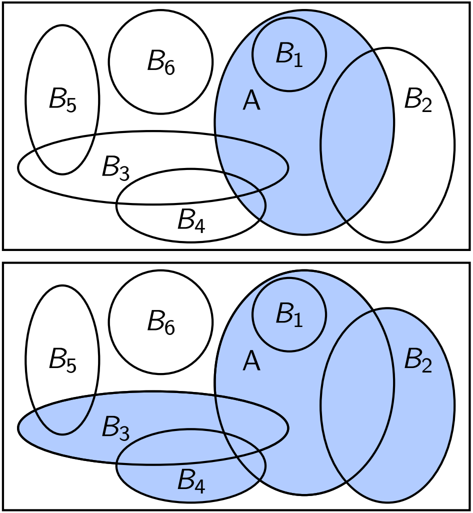
$$bel(A) = \sum_{B \subseteq A} m(B),$$

Plausibility function

Potential degree of belief that *could be* given to A:

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

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



Credal transformations

Credal partition

	$m_{i\emptyset}$	$m_{i\omega_1}$	$m_{i\omega_2}$	$m_{i\Omega}$
○	0	0	1	0
□	0	1	0	0
□	0	0.9	0.1	0
D	0	0	0	1
☆	1	0	0	0

plausibility →

Possibilistic partition, $\rightsquigarrow \overline{\lambda}$

	t_{i1}	t_{i2}
○	0	1
□	1	0
□	0.9	0.1
D	1	1
☆	0	0

belief →

$\rightsquigarrow \underline{\lambda}$

	t_{i1}	t_{i2}
○	0	1
□	1	0
□	0.9	0.1
D	0	0
☆	0	0

Credal transformations

Making decision : the pignistic transformation

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

Credal partition

	$m_{i\emptyset}$	$m_{i\omega_1}$	$m_{i\omega_2}$	$m_{i\Omega}$
○	0	0	1	0
□	0	1	0	0
▢	0	0.9	0.1	0
◗	0	0	0	1
☆	1	0	0	0

pignistic → transformation

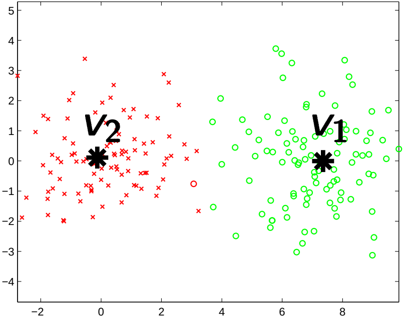
Fuzzy partition

	$u_{i\omega_1}$	$u_{i\omega_2}$
○	0	1
□	1	0
▢	0.9	0.1
◗	0.5	0.5
☆	0.5	0.5

Outline

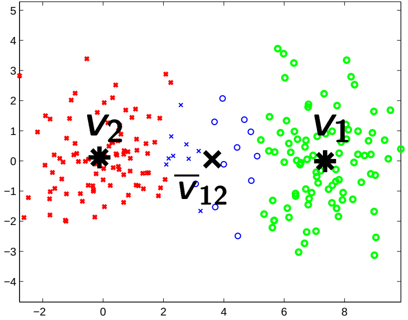
Evidential c-means (ECM)

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



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 $A_j \subseteq \Omega$
- Distance d_{ij}^2 between $\mathbf{x}_i$ and $\overline{\mathbf{v}}_j$



Evidential c-means (ECM)

Objective function

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

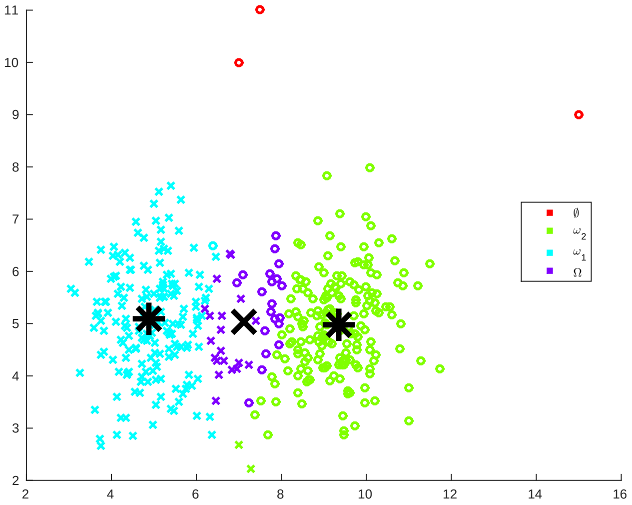
Subject to

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

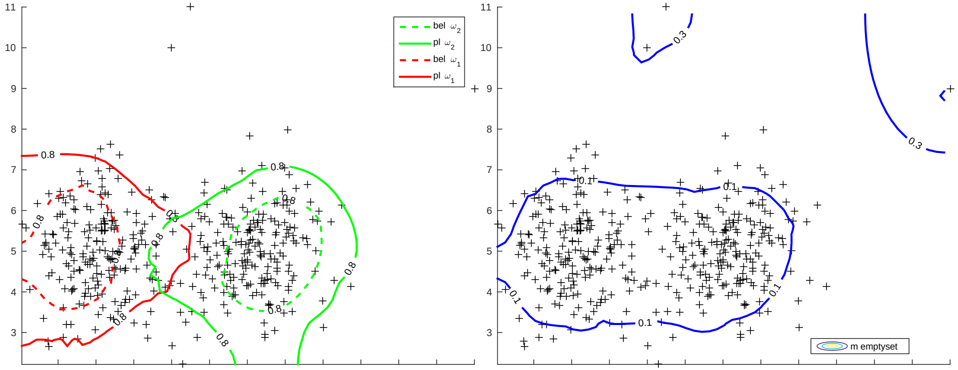
Alternate optimization

$$\text{opt}(\mathbf{M}) \leftrightarrows \text{opt}(\mathbf{V})$$

ECM: hard credal partition

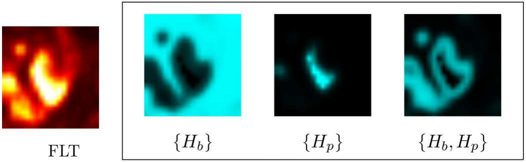


ECM: lower, upper bound and outliers

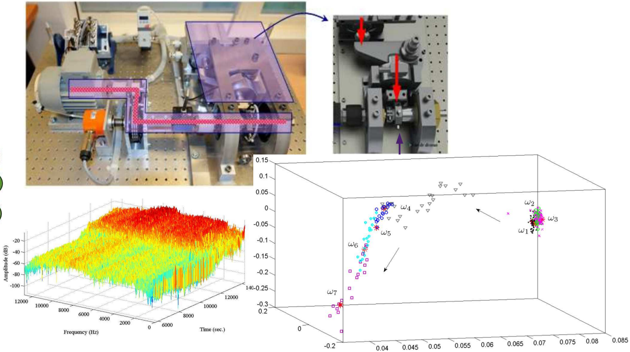


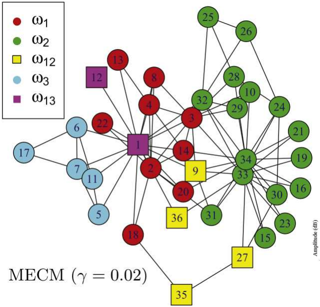
Applications

- medical image processing [3]
- machine prognosis [9]
- analysis of social networks [12]



Belief mass estimation

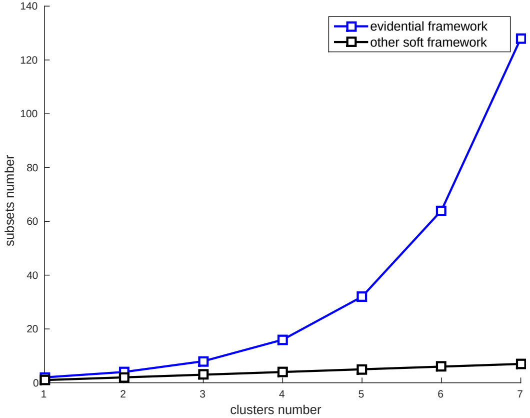




MECM ($\gamma = 0.02$)

Problematic

c clusters $\Rightarrow 2^c$ subsets !

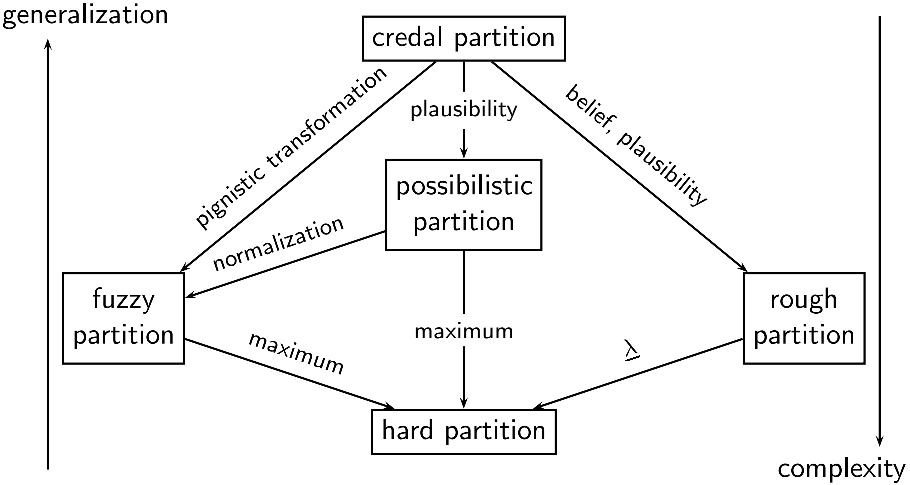


Outline

Clustering

- *k-means*
  - + real time constraint, big data, hard decision to make
- *fuzzy c-means*
  - + handle overlapped clusters
- *rough k-means*
  - + hard decision to make with overlapped clusters, upper and lower belief on that decision
- *possibilistic c-means*
  - + separate analysis of each cluster results, outliers
  - - mathematical framework with properties complex to handle
- *evidential c-means*
  - + strong partition analysis possible
  - - small number of clusters

Partition types



References I

V. Barra, V. Delouille, M. Kretzschmar, and J. Hochedez.
Fast and robust segmentation of solar euv images: algorithm and results for solar cycle 23.
Astronomy & Astrophysics, 505(1):361–371, 2009.

A. Chidambaram, H. Fillmore, T. Van Meter, C. Dumur, and W. Broaddus.
Novel report of expression and function of cd97 in malignant gliomas: correlation with wilms tumor 1 expression and glioma cell invasiveness.
Journal of neurosurgery, 116(4):843–853, 2012.

B. Lelandais, S. Ruan, T. Denœux, P. Vera, and I. Gardin.
Fusion of multi-tracer pet images for dose painting.
Medical image analysis, 18(7):1247–1259, 2014.

P. Lingras and C. West.
Interval set clustering of web users with rough k-means.
Journal of Intelligent Information Systems, 23(1):5–16, 2004.

P. Maji and S. Pal.
Rough set based generalized fuzzy *c*-means algorithm and quantitative indices.
IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 37(6):1529–1540, 2007.

G. Münz, S. Li, and G. Carle.
Traffic anomaly detection using k-means clustering.
In *GI/ITG Workshop MMBnet*, pages 13–14, 2007.

G. Peters.
Some refinements of rough k-means clustering.
Pattern Recognition, 39(8):1481–1491, 2006.

References II

P. Rizzoli, M. Martone, H. Rott, and A. Moreira.
Characterization of snow facies on the greenland ice sheet observed by tandem-x interferometric sar data.
Remote Sensing, 9(4):315, 2017.

L. Serir, E. Ramasso, and N. Zerhouni.
Evidential evolving gustafson-kessel algorithm for online data streams partitioning using belief function theory.
International Journal of Approximate Reasoning., 53(5):747–768, 2012.

O. Vincent, A. Makinde, O. Salako, and O. Oluwafemi.
A self-adaptive k-means classifier for business incentive in a fashion design environment.
Applied computing and informatics, 14(1):88–97, 2018.

A. Zare, N. Young, D. Suen, T. Nabelek, A. Galusha, and J. Keller.
Possibilistic fuzzy local information c-means for sonar image segmentation.
In *Computational Intelligence (SSCI), 2017 IEEE Symposium Series on*, pages 1–8. IEEE, 2017.

K. Zhou, A. Martin, Q. Pan, and Z. Liu.
Median evidential c-means algorithm and its application to community detection.
Knowledge-Based Systems, 74:69–88, 2015.

Thank you