

Fuzzy Inspirations in Unsupervised Learning

**In Depth Study for Uncertain Systems
February 2025**

Oltolini Edoardo - 869124



Fuzzy Sets in Data Analysis: From Statistical Foundations to Machine Learning

Inés Couso

Department of Statistics and OR, University of Oviedo, Gijón, SPAIN

Christian Borgelt

Department of Computer and Information Science, University of Konstanz, Konstanz, GERMANY

Eyke Hüllermeier

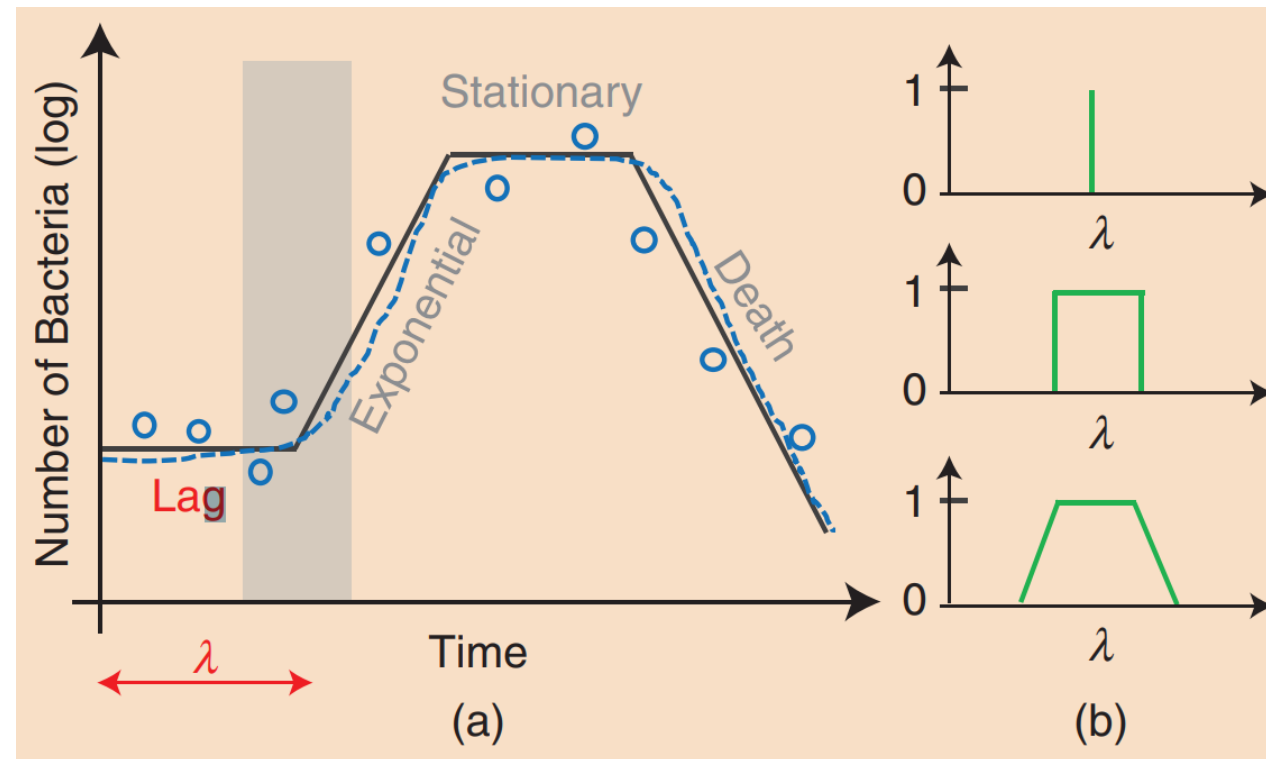
Department of Computer Science, Paderborn University, GERMANY

Rudolf Kruse

Faculty of Computer Science, Otto von Guericke University, Magdeburg, GERMANY

Perché l'incertezza è importante?

Nella Data Analysis, di solito si assume che i dati provengano da misurazioni precise, questa supposizione è solitamente non corretta.



Punto di Vista Epistemico vs Ontico

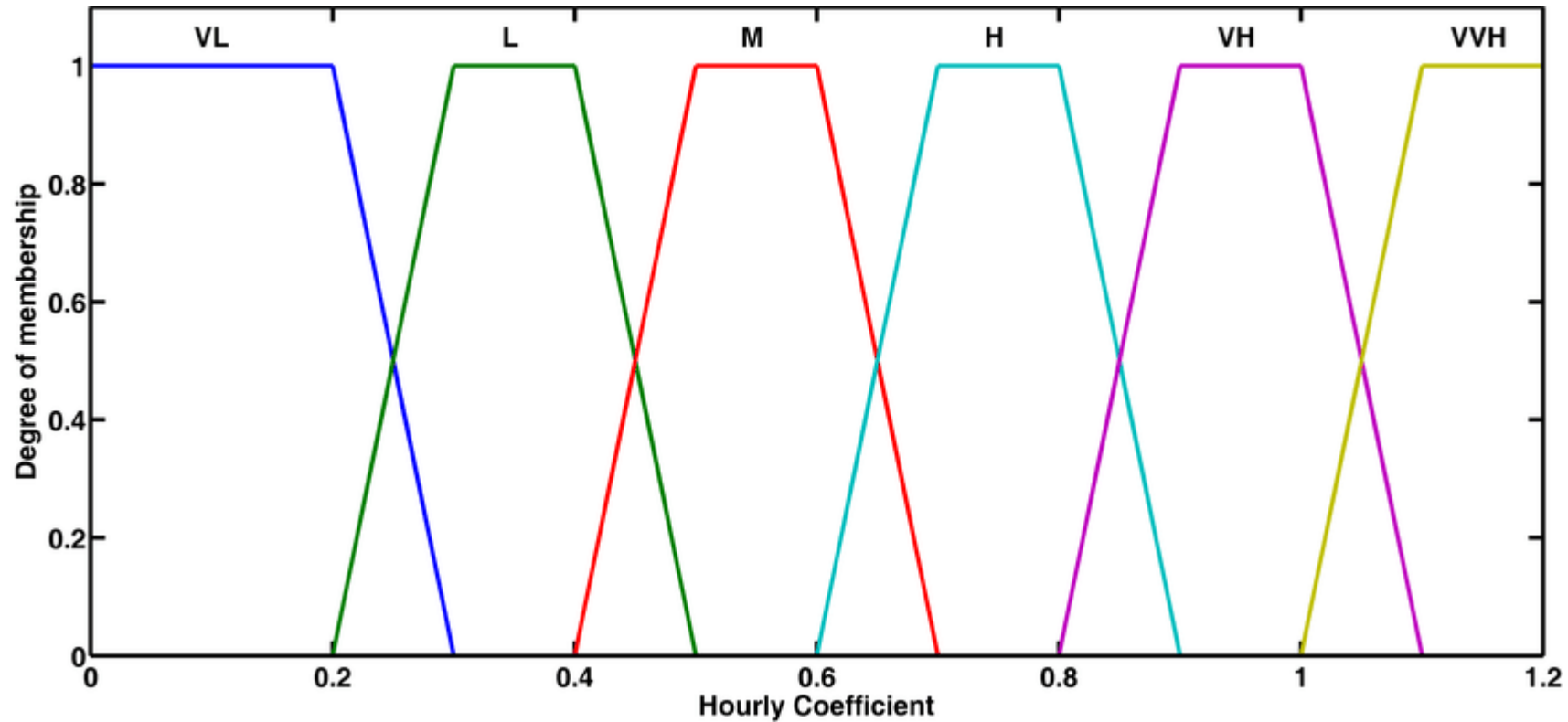
Ontico

Le entità sono punti nello spazio $F(X)$.
Il “Lag Time” è una quantità che può assumere valori di appartenenza a diversi Fuzzy Set: breve, medio, lungo.

Epistemico

La conoscenza su un oggetto sconosciuto $x_0 \in X$ è modellata da un Fuzzy Set \tilde{A} : $\forall x \in X$, $\tilde{A}(x)$ è il grado di possibilità che $x_0 = x$.

Punto di Vista Epistemico vs Ontico



Unsupervised Learning



Clustering: insieme di tecniche di analisi multivariata dei dati volte al raggruppamento di elementi in un insieme di dati.

Clustering Fuzzy?

Ontico

Fuzzy C-Means

Epistemico

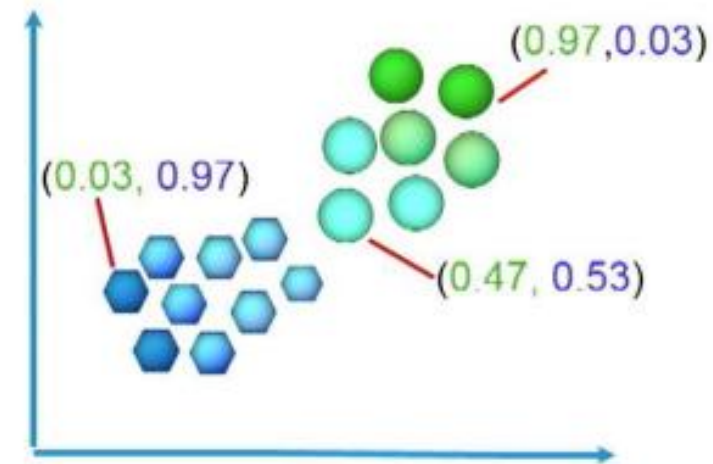
Possibilistic Clustering

Fuzzy Clustering

Hard Clustering



Soft Clustering



Gli algoritmi di fuzzy clustering, come il Fuzzy C-Means, sono del tipo «Soft Clustering», e assegnano ad ogni data point diversi gradi di appartenenza a ciascun cluster.

Fuzzy C-Means (FCM)

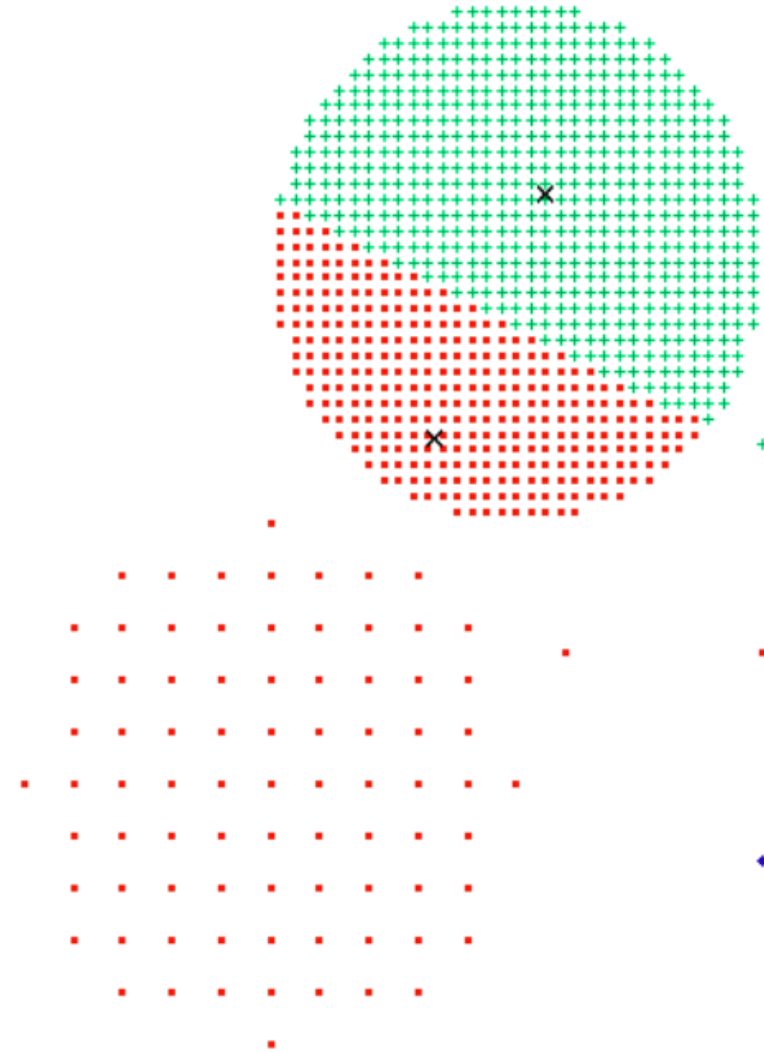
Ogni cluster C_i è rappresentato da un fuzzy set f_{C_i} .

$$\sum_i f_{C_i}(o_j) = 1.$$

Si assegna l'oggetto ai cluster con un grado di membership basato sulla distanza rispetto ad essi.

Vantaggi: cerca di evitare minimi locali.

Problemi: può generare minimi globali indesiderati. Problema degli outlier



Fuzzy C-Means

$$SSE_f = \sum_{i=1}^C \sum_{k=1}^N u_{ki}^w * d(x_k, v_i)^2$$

- u_{ij}^w è una funzione convessa in $[0, 1]$.
- $w > 1$, è detto “Fuzzifier”, esprime l’impatto dell’aggiunta di u_{ki} al K-Means

Fuzzy C-Means

- **Inizializzazione:** genero casualmente gli insiemi fuzzy (cioè i gradi/pesi “u”)
- **Iterazione:**
 - Aggiorno i centroidi per minimizzare SSE_f
 - Aggiorno i pesi per minimizzare SSE_f
- **Ripeto** fino al criterio di **convergenza** (cioè non diminuisce più l'errore) oppure ho superato un numero di iterazioni prefissato dall'utente.

$$u_{ki} = \frac{\frac{1}{D(x_k, v_i)^{\frac{1}{w-1}}}}{\sum_{j=1}^C \frac{1}{D(x_k, v_j)^{\frac{1}{w-1}}}} \quad v_i = \frac{\sum_{k=1}^N (u_{ki})^w * x_k}{\sum_{k=1}^N (u_{ki})^w}$$

Fuzzy Numbers, come scelgo w?

Un Sottoinsieme Fuzzy A di U è formato da elementi di U con un ordine gerarchico dato dai gradi di appartenenza. $x \in U$ sarà in una “order class” α se il suo “membership value” è $\geq \alpha$, con $\alpha \in [0, 1]$, reale.

$$[A]^\alpha = \{x \in U : \varphi_A(x) \geq \alpha\}, \text{ con } 0 \leq \alpha \leq 1, \text{ e } \varphi_A : \text{membership function}$$

Fuzzy Numbers

Un Fuzzy Set A è un numero fuzzy se l'Universo U su cui ϕ_A è definita, è l'insieme di tutti i numeri reali e soddisfa le seguenti condizioni:

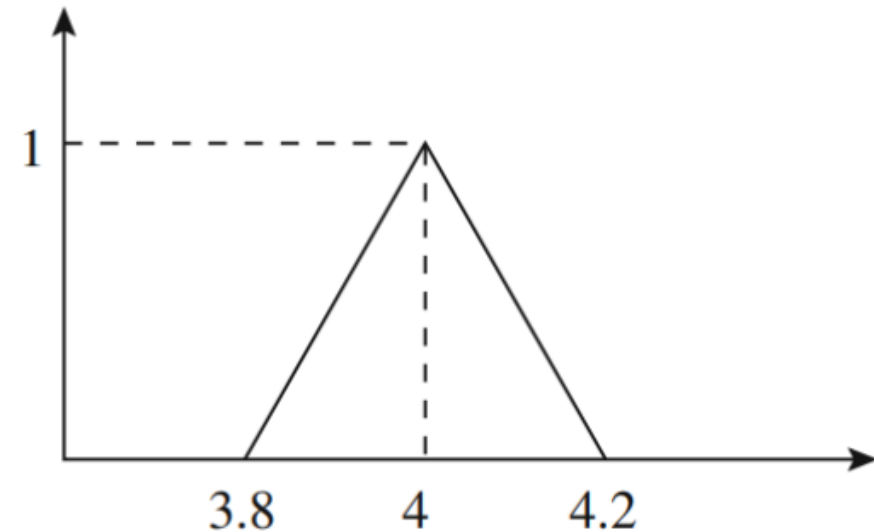
- (i) all the α -levels of A are not empty for $0 \leq \alpha \leq 1$;
- (ii) all the α -levels of A are closed intervals of \mathbb{R} ;
- (iii) $\text{supp } A = \{x \in \mathbb{R} : \varphi_A(x) > 0\}$ is bounded.

Fuzzy Number Triangolari

Di solito abbiamo a che fare con numeri fuzzy “triangolari”,
“trapezoidali” e “Gaussiani”.

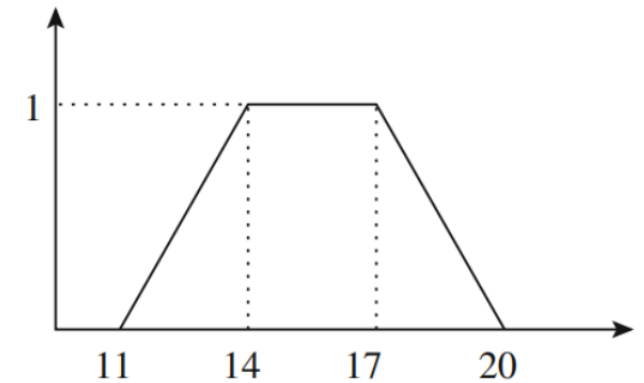
Un Fuzzy Set è “triangolare” se la sua funzione di membership è:

$$\varphi_A(x) = \begin{cases} 0 & \text{if } x \leq a \\ \frac{x-a}{u-a} & \text{if } a < x \leq u \\ \frac{x-b}{u-b} & \text{if } u < x \leq b \\ 0 & \text{if } x \geq b \end{cases}$$

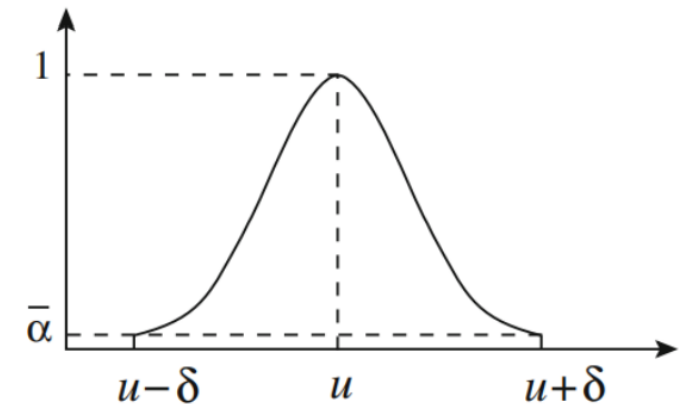


Trapezoidali e Gaussiani

$$\varphi_A(x) = \begin{cases} \frac{x-a}{b-a} & \text{if } a \leq x < b \\ 1 & \text{if } b \leq x \leq c \\ \frac{d-x}{d-c} & \text{if } c < x \leq d \\ 0 & \text{otherwise} \end{cases}$$

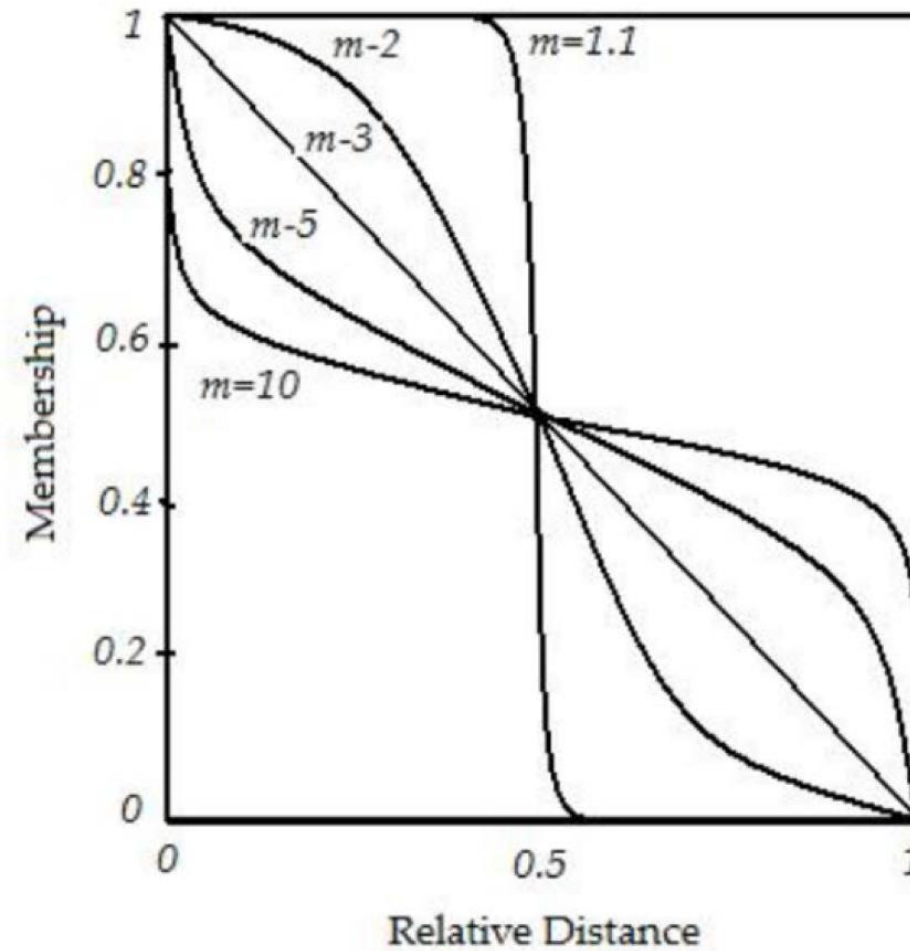


$$\varphi_A(x) = \begin{cases} \exp\left(-\left(\frac{x-u}{\alpha}\right)^2\right) & \text{if } u - \delta \leq x \leq u + \delta \\ 0 & \text{otherwise} \end{cases}$$

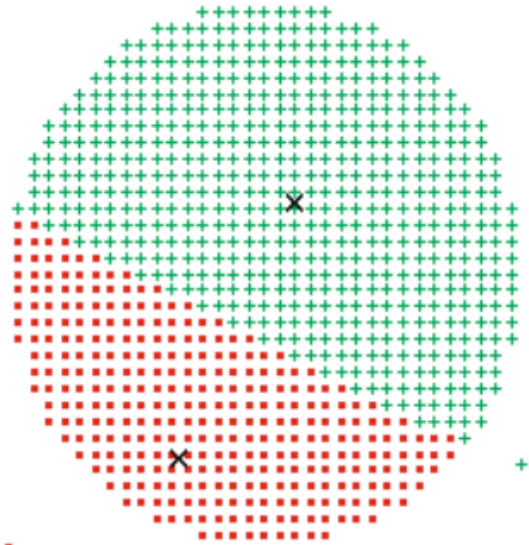


dove: $\bar{\alpha} = e^{-\left(\frac{\delta}{\alpha}\right)^2}$

FCM: Come scelgo m ?



Minimi Globali Indesiderati



Una delle possibili
soluzioni è
utilizzare un diverso
«fuzzifier»

Classico: $g(u) = u^w$



Basta che:

- $g(0) = 0$
- $g(1) = 1$

Deve essere facilmente
derivabile, Trasformazione
Quadratica?

Polynomial Fuzzifier

FCM – Polynomial Fuzzifier

$$f_{\text{PFCM}}(u) = \frac{1-\beta}{1+\beta}u^2 + \frac{2\beta}{1+\beta}u, \beta \in [0, 1)$$

$$J_{\text{PFCM}}(\mathbf{X}, \mathbf{U}, \mathbf{Y}) = \sum_{i=1}^c \sum_{j=1}^n f_{\text{PFCM}}(u_{ij}) d_{ij}^2$$

$d_{ij} > d_{kj} \rightarrow u_{ij} \leq u_{kj}$. I centroidi vengono ordinati in base alle loro distanze da x_j , ciò fornisce anche un ordinamento sui membership degree.

FCM – Polynomial Fuzzifier

$$u_{ij}^{t+1} \stackrel{\text{PFCM}}{=} \begin{cases} \frac{1}{1-\beta} \left(\frac{1 + (\hat{c}_j^t - 1)\beta}{\sum_{k=1}^{\hat{c}_j^t} \frac{(d_{ij}^t)^2}{(d_{\varphi(k)j}^t)^2}} - \beta \right) & \text{iff } \varphi(i) \leq \hat{c}_j^t \\ 0 & \text{otherwise} \end{cases}$$

FCM con Noise Cluster

Il rumore è quindi una classe separata, rappresentata da un centroide extra, che ha distanza costante δ , da tutti i vettori “feature”. La funzione da minimizzare diventa:

$$J_2(U, V) = \sum_{i=1}^c \sum_{k=1}^n (u_{ik})^m D_{ik} + \sum_{k=1}^n (u_{0k})^m \delta^2$$

Possibilistic Clustering → Teoria della Possibilità

«Tizio è **QUASI calvo**»: appartiene alle persone calve? - Membership Degree.

«Tizio è **QUASI CERTAMENTE calvo**» - Teoria della Possibilità



I Fuzzy Set sono funzioni $\pi : S \rightarrow [0,1]$

Se $\pi(x) = 1$, x è totalmente possibile.

Se $\pi(x) = 0$, allora è impossibile.

Esempio: un'automobile è scura → è possibile che sia blu, nera...
ma è impossibile che sia bianca.

Probabilità VS Possibilità

Totale ignoranza: $\forall x \in S, \pi(x) = 1$, tutti gli stati sono ugualmente e totalmente possibili.

Conoscenza completa: $\exists x \in S, \pi(x) = 1$ e $\forall b \neq x, \pi(b) = 0$.

La misura di Possibilità ha un suo duale, la misura di Necessità:

Misura di Possibilità: $\Pi(A) = \max_{a \in A} \{\pi(a)\}$.

Misura di Necessità: $N(A) = 1 - \Pi(A^C) = \min_{a \notin A} \{1 - \pi(a)\}$

Possibilistic Clustering

- Rilassa il vincolo sui gradi di membership. La somma ora può essere \leq o ≥ 1 .
Almeno un cluster deve contenere l'oggetto o_j , quindi sicuramente la somma è > 0 .
- Deve essere modificata la funzione obiettivo, altrimenti avremo sempre $u_{ki} = 0$
- Le f_{C_i} diventano distribuzioni possibilistiche, $f_{C_i}(o_j)$ rappresenta la **typicality** dell'oggetto o_j nel cluster C_i

Possibilistic: Funzione Obiettivo e Aggiornamenti

Nella versione di Krishnapuram e Keller, è formulata come:

$$\min_{\mathbf{U}, \mathbf{H}} J_{PKM} = \sum_{i=1}^n \sum_{g=1}^k u_{ig}^m d^2(\mathbf{x}_i, \mathbf{h}_g) + \sum_{g=1}^k \eta_g \sum_{i=1}^n (1 - u_{ig})^m,$$

s.t. $u_{ig} \in [0, 1], i = 1, \dots, n, \forall g = 1, \dots, k,$

Aggiornamenti:

$$u_{ig} = \frac{1}{1 + \left(\frac{d_{ig}^2}{\eta_g} \right)^{\frac{1}{m-1}}}.$$

$$\eta_i = K \frac{\sum_{j=1}^N u_{ij}^m d_{ij}^2}{\sum_{j=1}^N u_{ij}^m}.$$

Penalizzazione



Possibilistic Clustering – Repulsion Constraints:

$$\begin{aligned} \min_{\mathbf{U}, \mathbf{H}} J_{PkM-R} &= \sum_{i=1}^n \sum_{g=1}^k u_{ig}^m d^2(\mathbf{x}_i, \mathbf{h}_g) + \sum_{g=1}^k \eta_g \sum_{i=1}^n (1 - u_{ig})^m \\ &\quad + \sum_{g=1}^k \gamma_g \sum_{g'=1, g' \neq g}^k \frac{1}{\xi d^2(\mathbf{h}_g, \mathbf{h}_{g'})} \\ \text{s.t.} \quad &u_{ig} \in [0, 1], i = 1, \dots, n, \forall g = 1, \dots, k, \end{aligned}$$

Termine di Repulsione

Valutazione del Clustering

Silhouette: coesione e separazione

- Calcola a = distanza media di i dai punti nello stesso cluster
- Calcola b = $\min(\text{distanza media di } i \text{ dai punti in altri cluster})$
- Coefficiente di Silhouette: $s_i = (b - a) / \max(a, b)$

Come deve essere s_i ?

Fuzzy Silhouette

$$CS = \frac{1}{n} \sum_{j=1}^n S_j$$



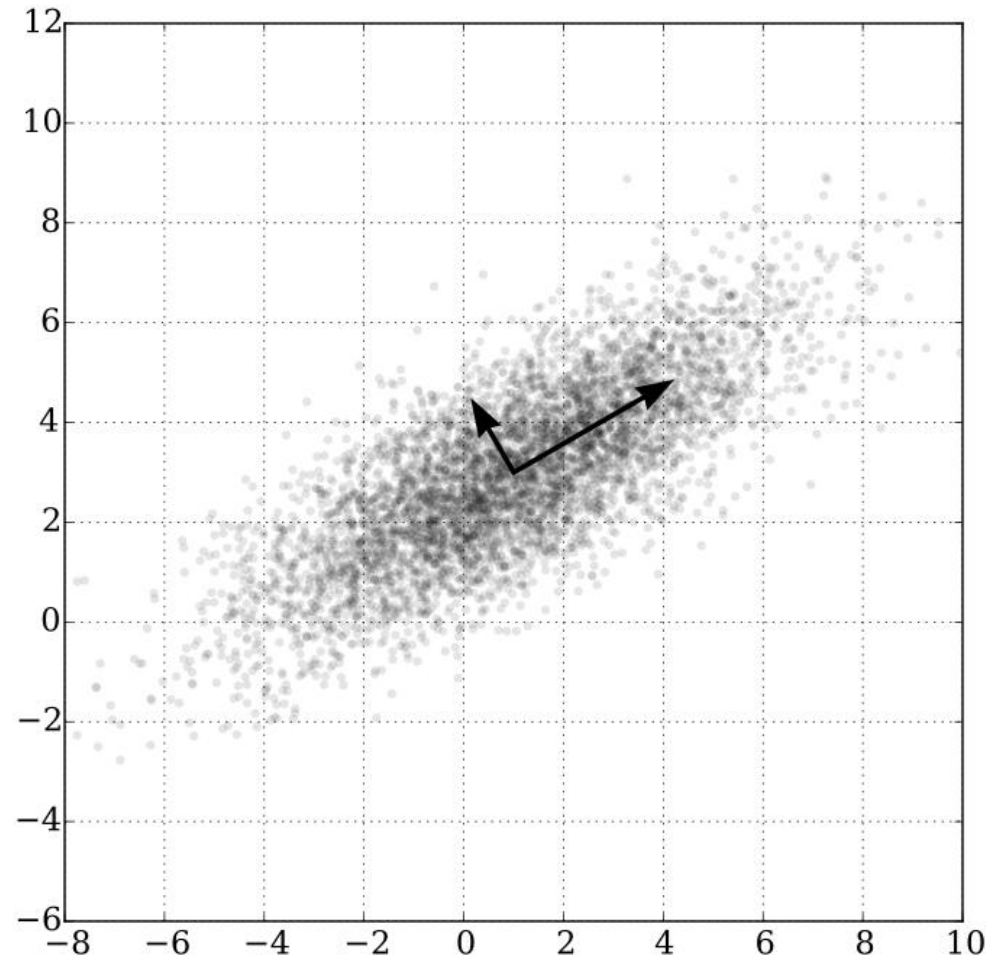
$$FS = \frac{\sum_{j=1}^n (u_{pj} - u_{qj})^{\alpha} S_j}{\sum_{j=1}^n (u_{pj} - u_{qj})^{\alpha}}$$

Xie-Beni Index: Compattezza e Separazione

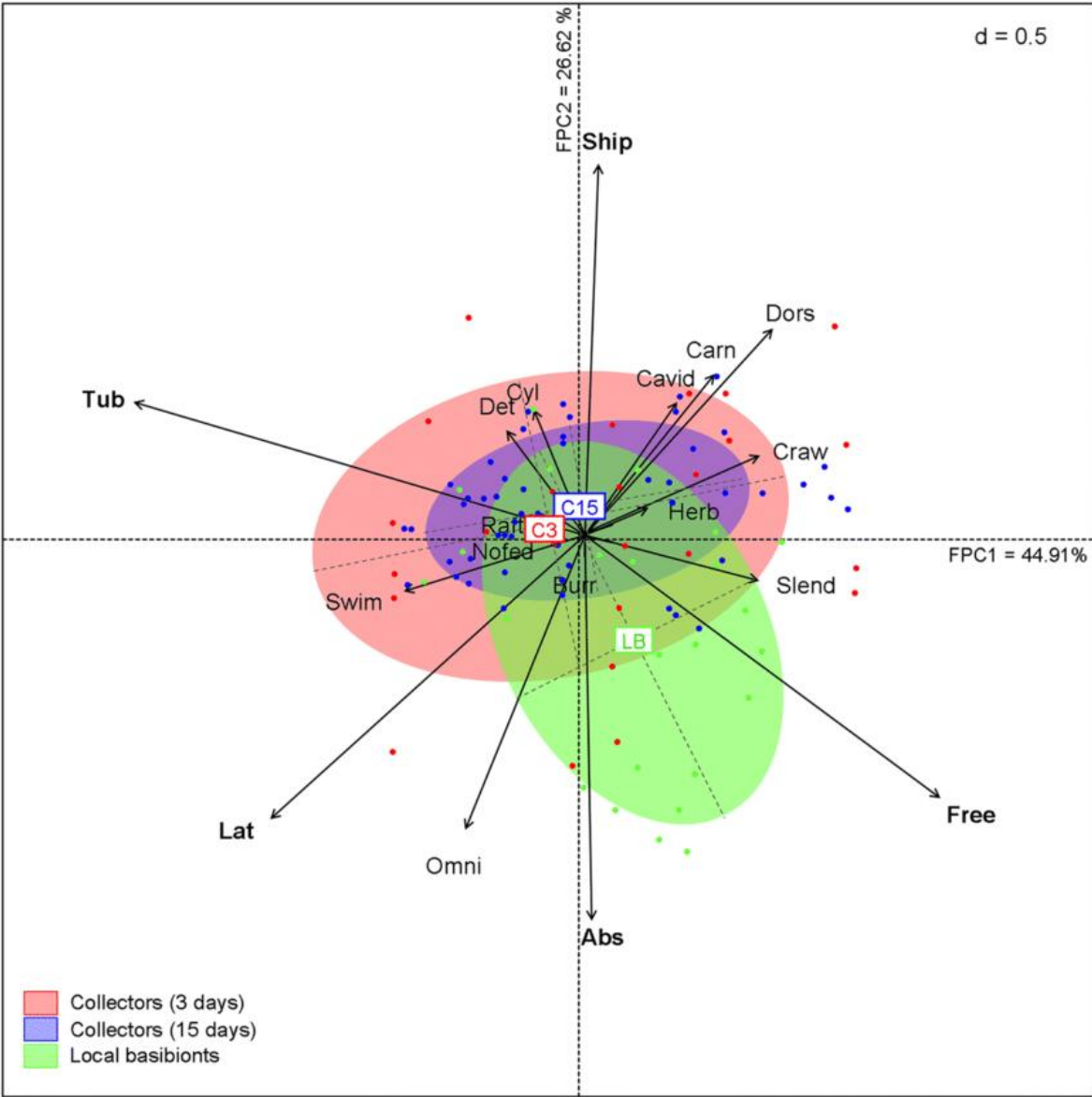
$$XB = \frac{\sum_{i=1}^N \sum_{k=1}^c u_{ik}^m \|x_i - v_k\|^2}{N \cdot \min_{j \neq k} \|v_j - v_k\|^2}$$

- N : numero di punti dati;
- c : numero di cluster;

Principal Component Analysis (PCA)



Fuzzy PCA



Fuzzy PCA, Fuzzy 1-Line

$$J(A, L; \alpha) = \sum_{j=1}^n [A(x^j)]^2 d^2(x^j, L) + \sum_{j=1}^n [\bar{A}(x^j)]^2 \frac{\alpha}{1 - \alpha}$$

$$J_2(U, V) = \sum_{i=1}^c \sum_{k=1}^n (u_{ik})^m D_{ik} + \sum_{k=1}^n (u_{0k})^m \delta^2$$

Robust fuzzy principal component analysis (FPCA). A comparative study concerning interaction of carbon-hydrogen bonds with molybdenum-oxo bonds

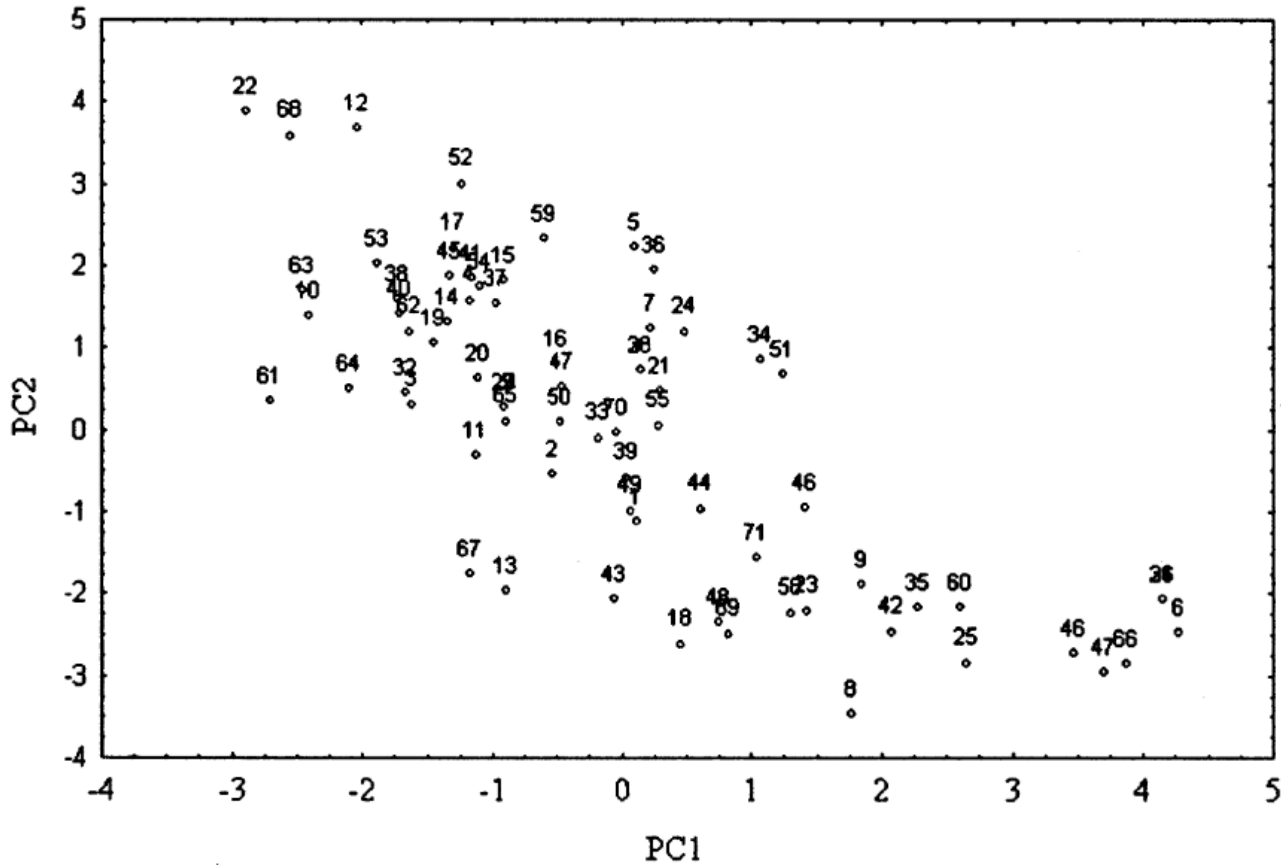
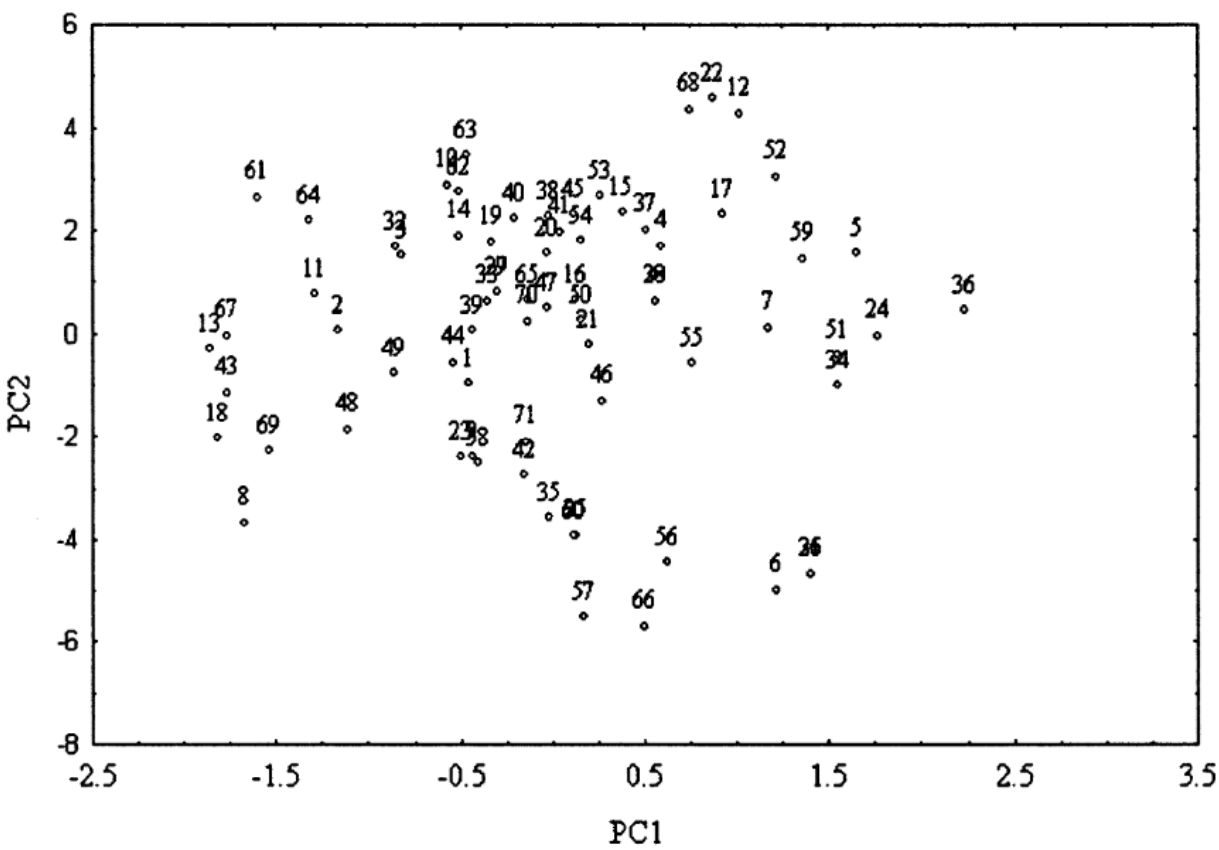
Fuzzy PCA

Il prototipo, la prima componente principale, viene denotata con $L(u,v)$, dove v è il centro della classe e u , con $\|u\| = 1$, che è la direzione principale, associata all'autovalore di modulo massimo λ_{max} della matrice di covarianza fuzzy:

$$C_{kl} = \frac{\sum_{j=1}^n [A_i(x^j)]^2 (x_{jk} - \bar{x}_k)(x_{jl} - \bar{x}_l)}{\sum_{j=1}^n [A_i(x^j)]^2}$$

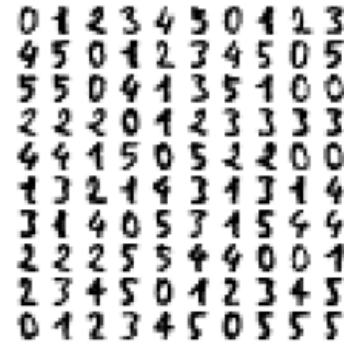
$$A(x^j) = \frac{\alpha/(1 - \alpha)}{[\alpha/(1 - \alpha)] + d^2(x^j, L)}$$

PCA vs FPCA



Digits Dataset

Shape: 1797x64



- **Contenuto:** 1797 immagini di cifre scritte a mano.
- **Formato:** Ogni immagine è una matrice 8x8 → 64 caratteristiche.
- **Scala di grigi:** I valori dei pixel vanno da 0 (bianco) a 16 (nero).
- **Etichette:** Ogni immagine ha un'etichetta che indica la sua cifra (da 0 a 9).

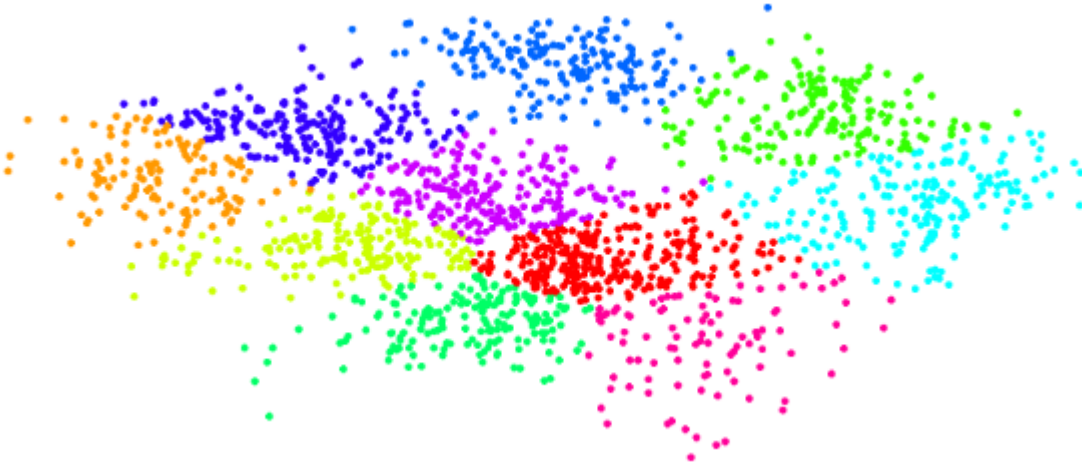
Metriche	FCM-PF	KM	FCM
Silhouette (Sil)	0.2984	0.1687	-0.0584
Silhouette con Fuzzy (SilF)	0.3755	—	0.0417
Xie-Beni (XB)	2.2441	1.6271	6.0807e+26

Digits + FPCA, n = 2

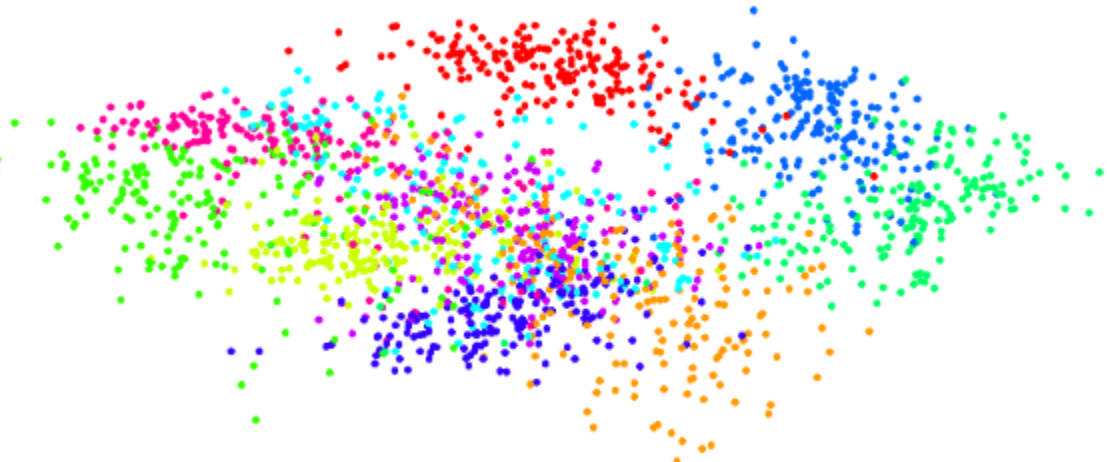
Varianza Spiegata	PC1 + PC2
FPCA	0.6883
PCA	0.2851

Metriche	FCM-PF	PCM	KM	FCM
Silhouette (Sil)	0.5641	-0.3541	0.3844	0.5559
Silhouette con Fuzzy (SilF)	0.6239	0.0473	—	0.6904
Xie-Beni (XB)	0.2101	374000.1748	0.2785	0.1445

FCM Algorithm

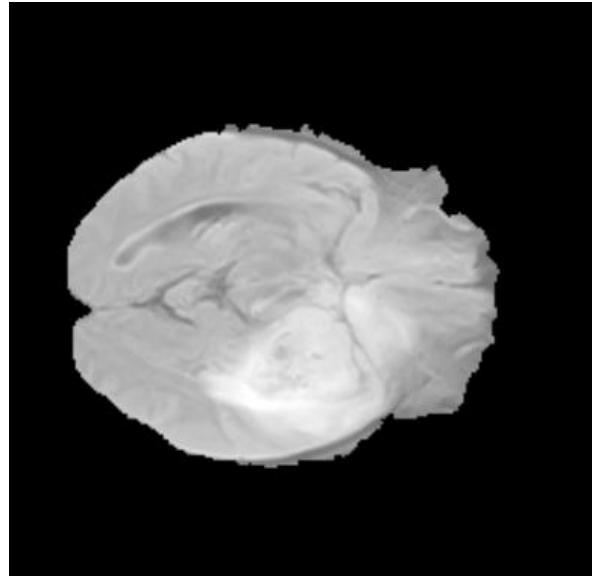


Labeled Data



FCM in Brain Tumor Segmentation

I diversi tessuti cerebrali, tra cui il liquido cerebrospinale, la materia bianca e la materia grigia, non sono ben distinguibili nelle RM, di conseguenza, i bordi tra le regioni si confondono tra loro, come nella seguente sezione:



FCM – Gustafson-Kessel

$$\mathbf{F}_i = \frac{\sum_{j=1}^N \mu_{ij}^m (\mathbf{x}_j - \mathbf{c}_i)(\mathbf{x}_j - \mathbf{c}_i)^\top}{S_i}, \quad 1 \leq i \leq C$$

$$S_i = \sum_{j=1}^N \mu_{ij}^m$$

$$D_{ij} = \sqrt{(\mathbf{x}_j - \mathbf{c}_i)^\top [\det(\mathbf{F}_i)^{1/N} \mathbf{F}_i^{-1}] (\mathbf{x}_j - \mathbf{c}_i)}, \quad 1 \leq i \leq C, \quad 1 \leq j \leq N$$

Brain Tumor Segmentation con FCM (GK)

Il processo di segmentazione include i seguenti passaggi:

1. Preparare i vettori delle caratteristiche dai dati di addestramento e di test.
2. Raggruppare i dati di addestramento utilizzando la metrica di distanza specificata.
3. Identificare il cluster che rappresenta il tumore.
4. Utilizzare i centri dei cluster identificati per individuare i tumori nei dati di test.

<https://it.mathworks.com/help/fuzzy/brain-tumor-segmentation-using-fuzzy-c-means-clustering.html>

Brain Tumor Segmentation con FCM (GK)

Questo esempio utilizza il dataset BraTS, che contiene volumi 4D. Ogni volume è di dimensioni $240 \times 240 \times 152 \times 4$. Le prime tre corrispondono all'altezza, larghezza e profondità dell'immagine 3D, mentre la quarta rappresenta le diverse modalità di scansione.

```
trainDataFileName = fullfile(imageDir,...  
    "sampleBraTSTestSetValid","imagesTest","BraTS447.mat");  
testDataFileName = fullfile(imageDir,...  
    "sampleBraTSTestSetValid","imagesTest","BraTS463.mat");  
testLabelFileName = fullfile(imageDir,...  
    "sampleBraTSTestSetValid","labelsTest","BraTS463.mat");
```

Brain Tumor Segmentation con FCM (GK)

```
kDim = [3 3];  
trainFeatures = createMovingWindowFeatures(trainingData,kDim);
```

```
refTumor = cell(1,n); % Tumor pixel ids  
refHasTumor = false(1,n);  
for id = 1:n  
    refTumor{id} = find(refLabel(:, :, id) == 1);  
    refHasTumor(id) = ~isempty(refTumor{id});  
end
```

Brain Tumor Segmentation with FCM (GK)

```
rng("default")
mnCenters = fcm(trainFeatures, options);

mnDist = findDistance(mnCenters, testFeatures);

[~, mnLabel] = min(mnDist', [], 2);

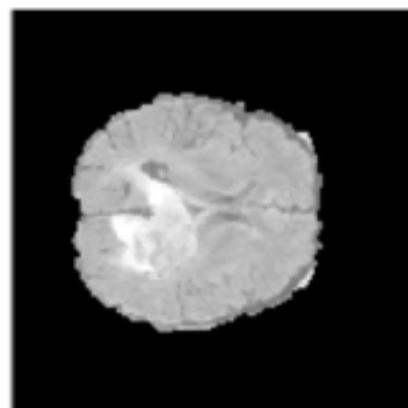
mnLabel = reshape(mnLabel, n, r*c)';
mnLabel = reshape(mnLabel, [r c n]);
```


Brain Tumor Segmentation with FCM (GK)

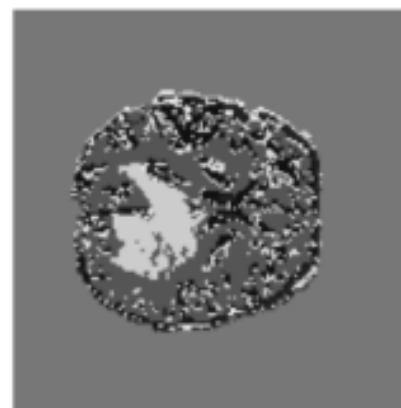
Ora identifichiamo i pixel tumorali nelle immagini di test con la funzione «segmentTumor», specificando l'indice del cluster che rappresenta il tumore.

- Etichetta i pixel tumorali nell'immagine.
 - Calcola il numero di pixel falsi positivi.
 - Restituisce un valore logico che indica se l'immagine contiene un tumore.
-

Segmentation with FCM (GK)



Test Image



Segmented Image



Tumor Detection



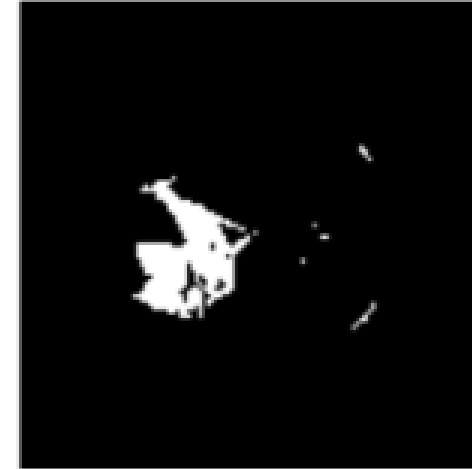
Labeled Image



Ground Truth



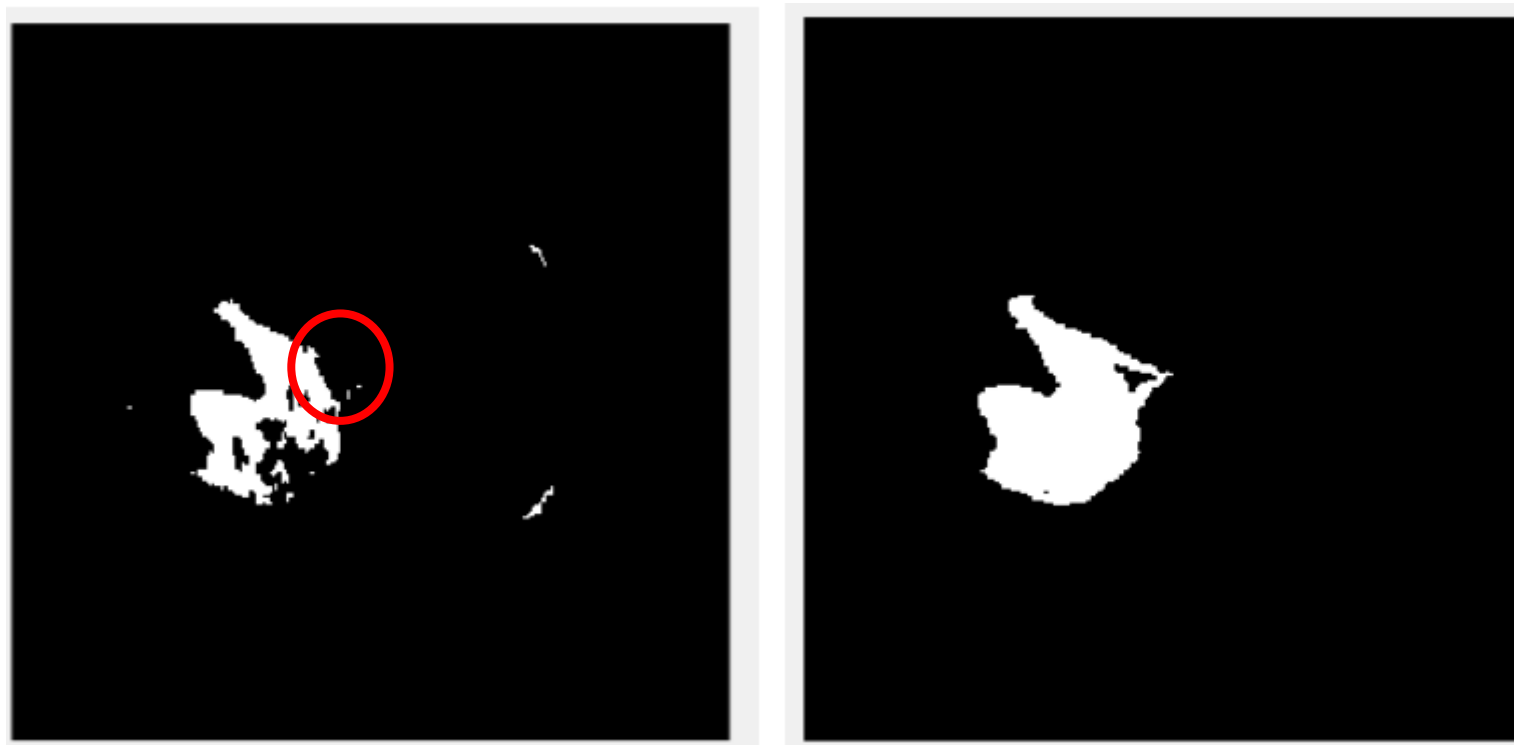
Euclidean Distance



Mahalanobis Distance

	True Pos	True Neg	False Pos	False Neg
Euclidean Distance	80	26	46	0
Mahalanobis Distance	79	36	36	1

Confronto con K-Means



Conclusioni

1. Prestazioni Migliorate:

I risultati dimostrano che il fuzzy clustering supera il k-means standard in termini di qualità, essendo più robusto e cercando minimi globali.

2. Segmentazione Ottimizzata dei Tumori Cerebrali:

L'FCM-GK riesce a identificare con maggiore precisione le boundary region, dove le caratteristiche dei tessuti tumorali e sani si sovrappongono, migliorando l'accuratezza nella segmentazione.

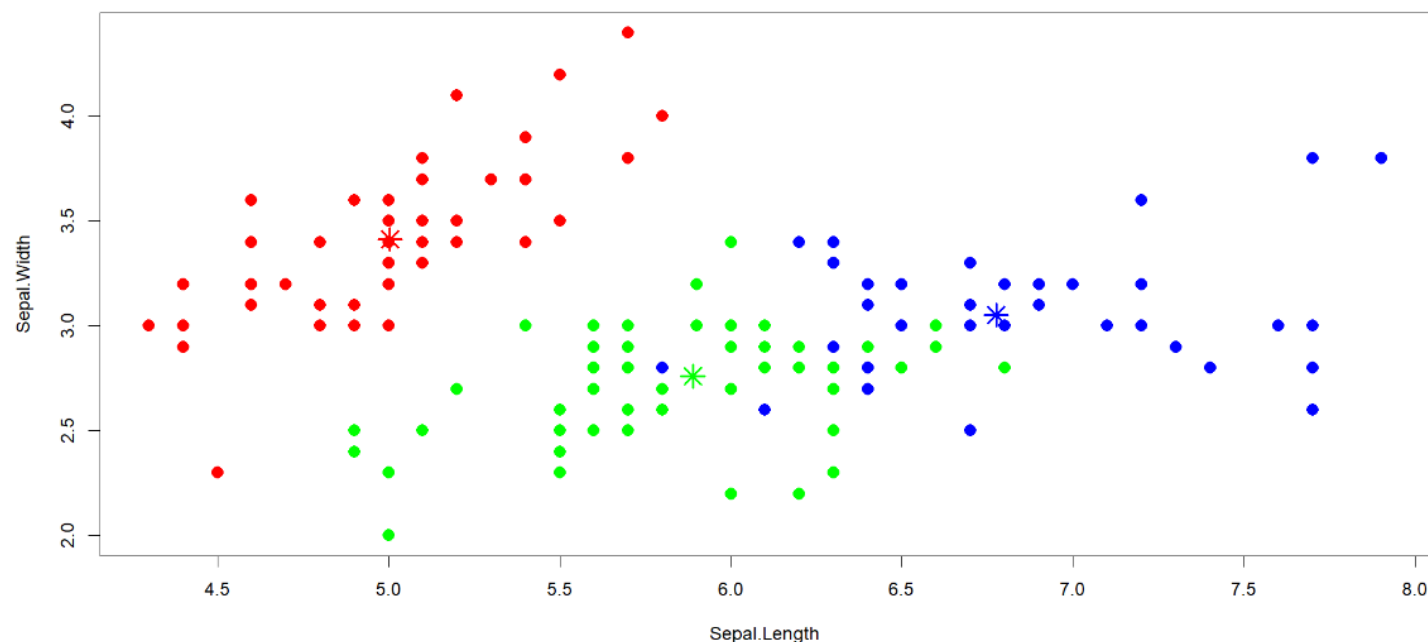
Risultati Finali con FCM (GK)

	True Pos	True Neg	False Pos	False Neg
	<hr/>	<hr/>	<hr/>	<hr/>
Euclidean Distance	80	26	46	0
Mahalanobis Distance	79	36	36	1

Calcolo delle Componenti Connesse + Rimozione Elementi Piccoli

True Pos	True Neg	False Pos	False Neg
<hr/>	<hr/>	<hr/>	<hr/>
79	61	11	1

Fuzzy Clustering su Iris Dataset



Metriche	FCM-PF	KM	FCM
Silhouette (Sil)	0.7357	0.5528	0.7320
Silhouette Fuzzy (SilF)	0.7578	—	0.8091
Xie-Beni (XB)	0.1543	0.1628	0.1369

Iris + FPCA, n=2, varianza 100%

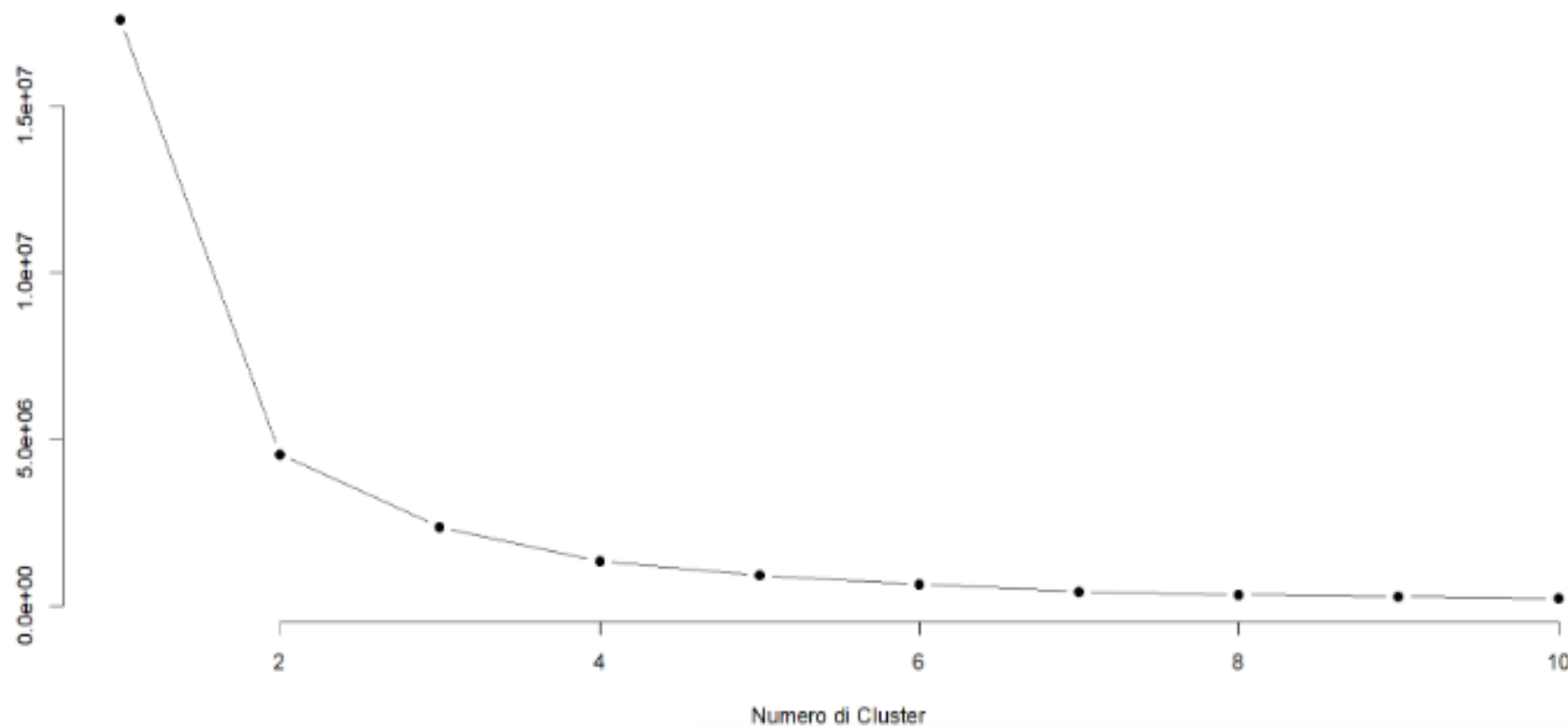
Metriche	FCM-PF	KM	FCM
Silhouette (Sil)	0.8023	0.6056	0.6286
Silhouette con Fuzzy (SilF)	0.8061	—	0.7131
Xie-Beni (XB)	0.1564	0.1557	0.3534

Possibilistic + Repulsion

Metriche	Valore
Silhouette (Sil)	0.7380
Silhouette con Fuzzy (SilF)	0.7988
Xie-Beni (XB)	0.0741

Wine Dataset

Il dataset contiene informazioni su **178 campioni di vino** con **13 caratteristiche chimiche**.



Wine Dataset

Pre-PCA

Metriche	FCM-PF	KM	FCM
Silhouette (Sil)	0.7125	0.5422	0.7117
Silhouette con Fuzzy (SilF)	0.7383	–	0.7876
Xie-Beni (XB)	0.1904	0.2077	0.1393

Table 8: Risultati delle metriche per PF, KM e FCM.

Post-PCA
n=2

Metriche	FCM-PF	PCM	KM	FCM
Silhouette (Sil)	0.6740	0.0247	0.4811	0.6698
Silhouette con Fuzzy (SilF)	0.7130	0.5718	–	0.7718
Xie-Beni (XB)	0.1114	467.8961	0.1643	0.0875

Table 9: Risultati delle metriche per PF, PCM, KM e FCM.

Wine Dataset

