Me llamo Jim Bezdek

Hola amigos !

Vivo en Pensacola

Amberjack
Pensacola, FL


Paraiso (363 dias al año)
EXPLORING NEW FRONTIERS
FUZZ-IEEE 2003
Speckled Trout
Pensacola, FL

Los otros dos?

NO ES TAN CALIENTE !

Speckled Trout
Pensacola, FL

Antes de que el huracán Iván

DESPUÉS DEL HURACÁN IVÁN

Trabajo en la Universidad de Melbourne

Mudshark
Melbourne, Au

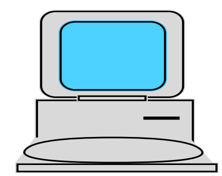
Wahoo and Ulua
Maui, Hawaii

Kawai, Napier, NZ

¿Qué son los grupos?

Para los humanos

Para las computadoras



Clusters from the *human point of view* (HPOV)

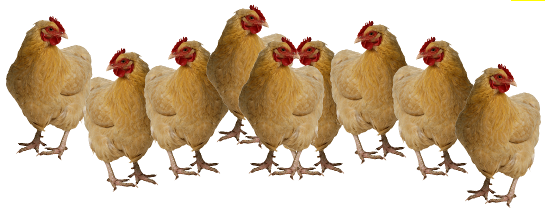
1st Google Hit	*a grouping of a number of similar things*
Encarta World Dictionary (v)	*to gather into or form a small group*
Online Free Dictionary (n)	*a* group *of the same or* similar *elements gathered or occurring* closely *together*
3 ingredients	*group* similar ≠ *close*

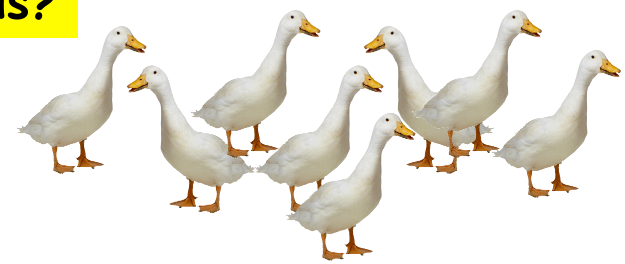
Similar (objects) ≠ close (objects) for humans

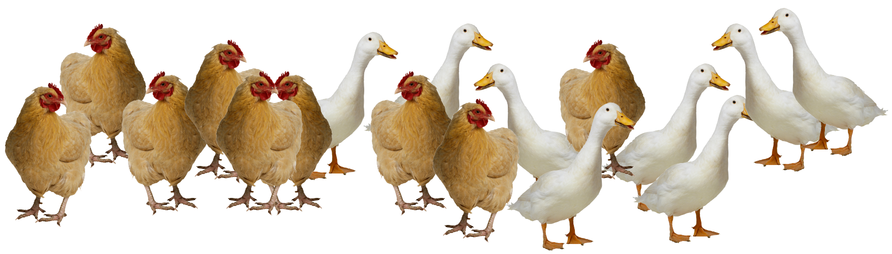
Similar (vectors) = (close) vectors for computers

Different observers may see similarity differently

2 clusters of birds?

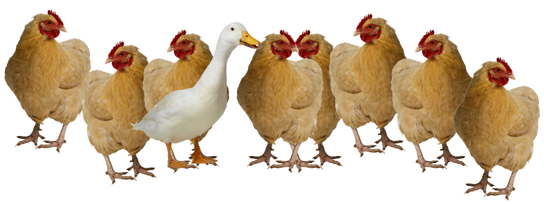






2 clusters of birds … or

1 cluster of birds with 2 kinds of physical labels?



1 cluster of birds … or

1 cluster of chickens
2 labels, 1 is noise ?

When do humans "see" *several* clusters?

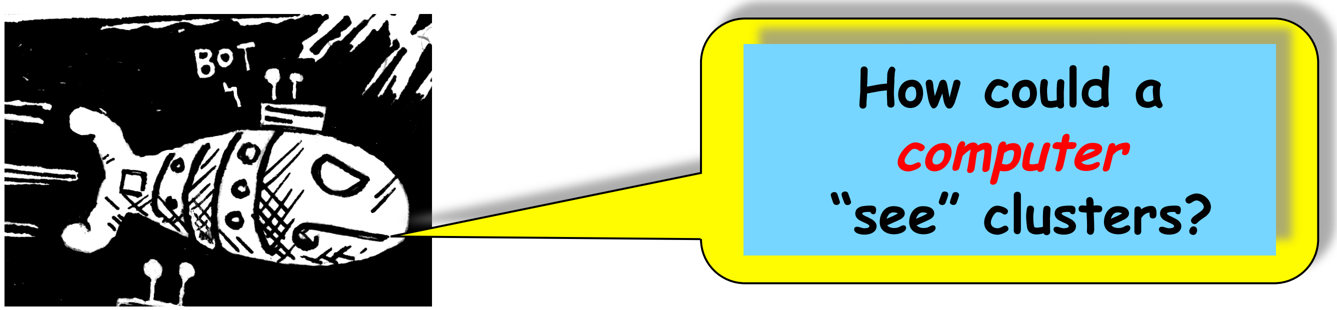
HHHHHHTTTTTTTTHHHHH	1, 2 or 3 ?
HHHHHTTTTTTTT HHHHH	1 or 2 ?
HHHHHH TTTTTTT HHHHH	Now?

1, 2 or 3 ?	HHHHHHHHHHTHHHHHHHHH
1 or 2 ?	HHHHHHHHHH THHHHHHHHH
Now ?	HHHHHHHHHH T HHHHHHHH

HTHTHTHT	1 or 2 ?
H T H T H T H T	1, 2 or 8 ?

When subsets have "enough" similarity, compactness and separation

So, humans *may* "see" clusters of real objects, but…



Computers use (Numerical) *data* | *Object Data* | *Relational Data*

Cluster analysis comprises computational models and algorithms that search for *groups* of *similar ("close")* data that are *"well separated"*

We hope that *computer point of view (CPOV)* clusters in *data* correspond to *HPOV clusters* in the *objects* represented by the data

[*CPOV)*] clusters are defined *ONLY by each model and algorithm*. So e.g., we can have HCM/AO ≠ HCM/PSO ≠ HCM/GA clusters

… and, we *NEVER* know *CPOV clusters* in the *objects* represented by almost all numerical data (only object data for $p \leq 3$)

2 Kinds of Data for Pattern Recognition

Objects

$O = \{o_1, \ldots, o_n\}$: o_i = i-th *physical* object

Object Data

$X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^p$: x_i = *feature vector* for o_i

x_{ji} = j-th (*measured*) feature of x_i : $1 \leq j \leq p$

Relational Data

$R = [r_{ij}]$ = *relationship* (o_i, o_j)) or (x_i, x_j)

d_{ij}

s_{ij} = pairwise *similarity* (o_i, o_j) or (x_i, x_j)

d_{ij} = pairwise *dissimilarity* (o_i, o_j) or (x_i, x_j)

Tyically

(R = D)

$$\left.\begin{array}{ll} d_{ii} = 0 & : 1 \leq i \leq n \\ d_{ij} > 0 & : 1 \leq i, j \leq n \end{array}\right\} \quad \text{(Positive-definite)}$$

$$d_{ij} = d_{ji} \quad : 1 \leq i \neq j \leq n \qquad \text{(Symmetric)}$$

We often convert X→D with *distance* $d_{ij} = \left\| x_i - x_j \right\|$

Cluster Analysis *in Numerical Data* : 3 Problems

Preclustering	Partitioning	Post-clustering
1. Assessment	**2. Clustering**	**3. Validation**
Does the data have (c) clusters?	Find clusters U(c) in X or R	U(c) ok? (useful, realistic, etc.)

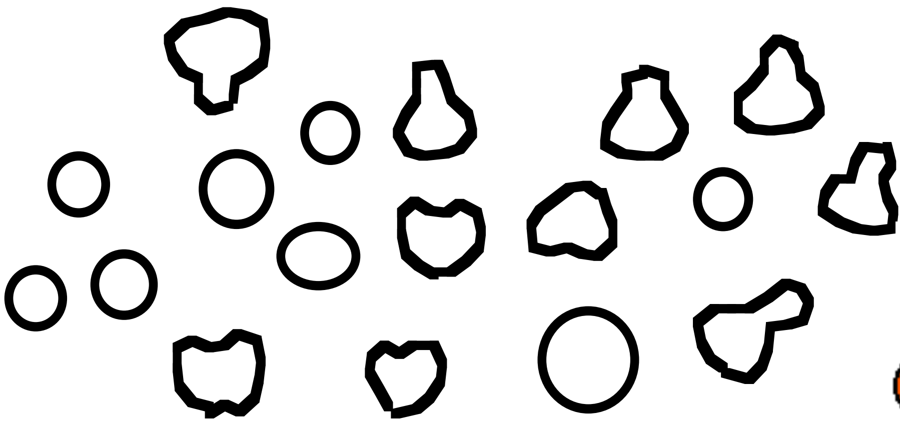
"Correlation profile analysis is a simple method of discovering and grouping together variables having identical patterns, …, which are said to comprise a cluster. The objective of [factor] analysis is to discover clusters and the arrangement of these."

Tryon, R. C. (1939). Cluster Analysis, Edwards Bros., Ann Arbor, MI

"Now we face the much nastier problem of determining the appropriate value for [c]."

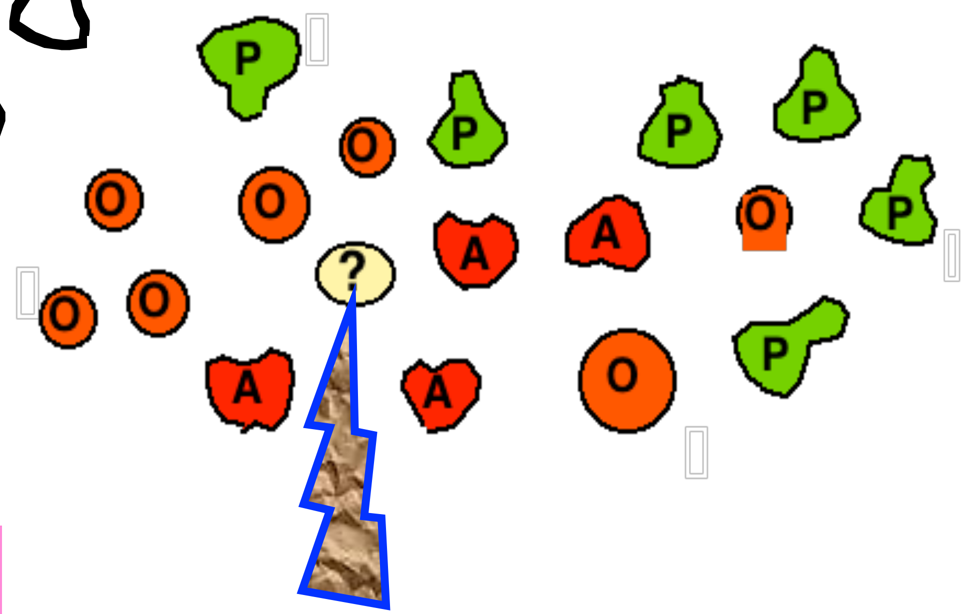
Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika*, 18(4), 267-276.

Label the similar objects



What is Clustering ?

(Unsupervised Learning}



What is "similar"?

What is "close"?

How many groups (c) ?

Easy for Humans !

Hard for computers !

Measuring (Dis)Similarity : The Usual Suspects

Mahalonobis

< , > Norms

Diagonal

$$d_{Maha} = \|x - v\|_{M^{-1}} = \sqrt{(x-v)^T M^{-1} (x-v)}$$

$$d_{diag} = \|x - v\|_{D_M^{-1}} = \sqrt{(x-v)^T D_M^{-1} (x-v)}$$

Euclidean

$$d_2 = \|x - v\|_I = \sqrt{(x-v)^T (x-v)}$$

City Block

Minkowski Norms

Sup or Max

$$d_1 = \|x - v\|_1 = \sum_{j=1}^{p} |x_j - v_j|$$

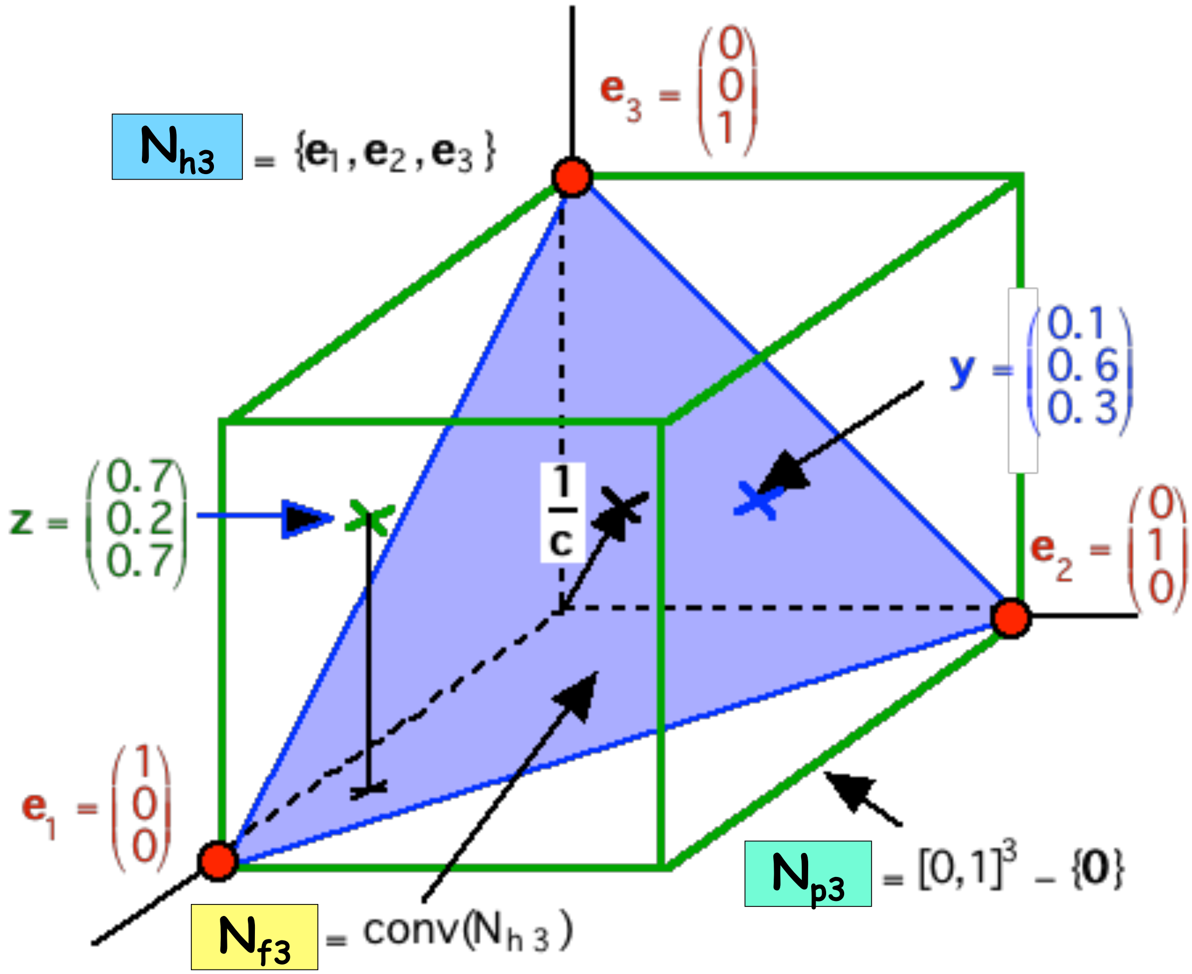
$$d_\infty = \|x - v\|_\infty = \max_{1 \le j \le p} \{|x_j - v_j|\}$$

Label Vectors @ c = 3

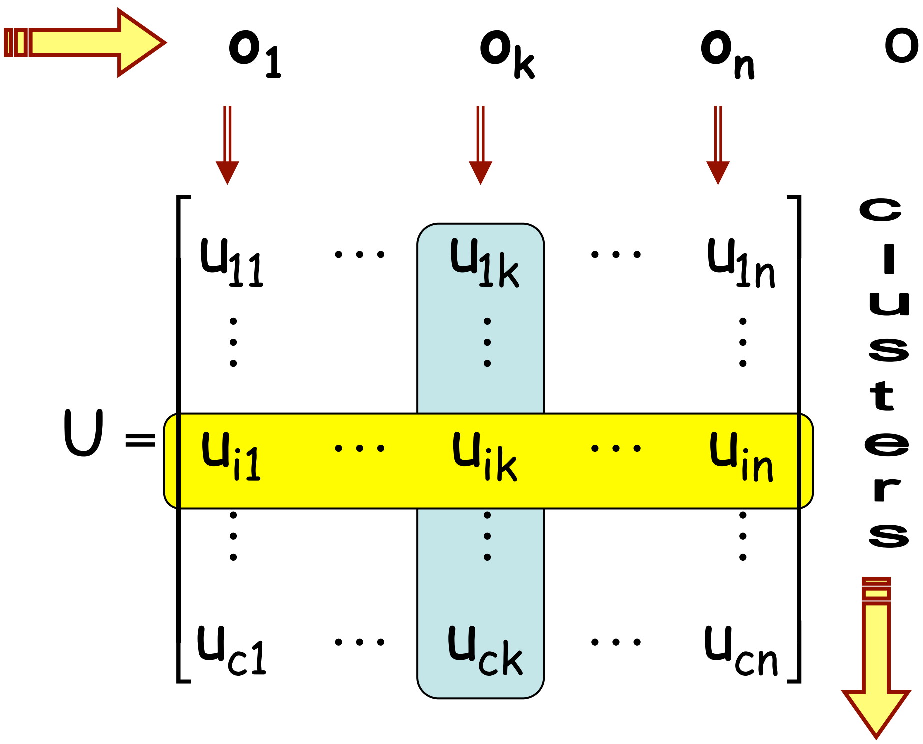
Crisp
= Vertices

Fuzzy or
Probabilistic
= Triangle

Possibilistic
= Cube-{θ}



row i ➠ M'ship of all o_k's in cluster i



col k ➠ M'ship of o_k in each cluster

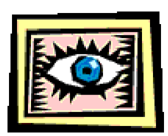
Partition Matrices

Membership Functions

$u_i : O \rightarrow [0,1]$

$u_i(o_k) = u_{ik}$ = M'ship of o_k in cluster i

	Crisp	Fuzzy/Prob	Possibilistic
Row sums	$\sum_{k} u_{ik} > 0$	same	same
Col sums	$\sum_{i} u_{ik} = 1$	same	$\sum_{i=1}^{c} u_{ik} \le c$
M'ships	$u_{ik} \in \{0,1\}$	$u_{ik} \in [0,1]$	same
Set Name	M_{hcn} $\subset$	M_{fcn} $\subset$	M_{pcn}
Example	1 0 0 0 0 1 0 0 0 0 1 1	1 .07 0 .44 0 .91 0 .06 0 .02 1 .50	1 .07 1 .44 0 .91 0 .52 0 .02 1 .38

Take a 2nd 

Hardening Partitions

Defuzzification (Deprobabilization = Bayes Rule for $U \in M_{fcn}$!)

Defusification – we *defusify* bombs, situations – blah blah …

$$U = \begin{bmatrix} 1 & .07 & .6 & .5 \\ 0 & .81 & .2 & 0 \\ 0 & .12 & .7 & .5 \end{bmatrix} \in M_{pcn} \qquad H_1(U) = U_{MM} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \in M_{hcn}$$

resolve ties randomly

α -cut thresholding ($0< \alpha <1$) filters chosen levels of m'ship

$$H_{.9}(U) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \in M_{hcn} \qquad H_{.8}(U) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \in M_{hcn}$$

3 Types of Basic Clustering Models

1. *U only models* : find *clusters U* $\in M_{pcn}$

- Usually *relational models*, e.g., single linkage
- V=F(U) can be computed - if you want it

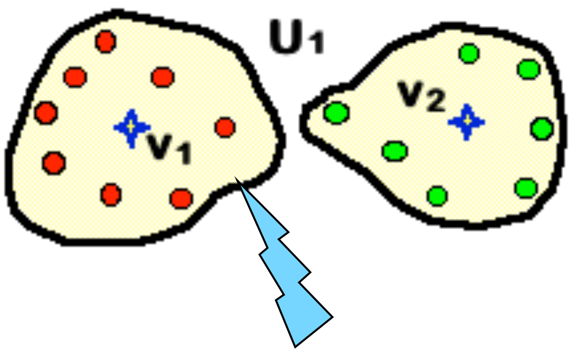
2. *V only models* : find (typically) *point prototypes V* $\in R^{cp}$

- Usually *local sequential* models, e.g., SOM
- U=G(V) can be computed - *super important* for VL Data Methods

***(U,V) models* : find *(U,V)* $\in M_{pcn} \times R^{cp}$**

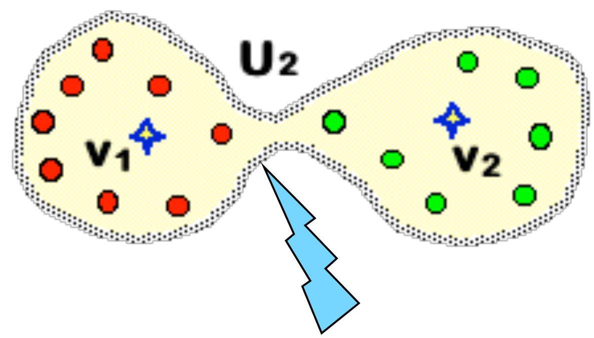
- 3A. Usually *global batch* models, e.g., c-Means
- 3B. (*U, V, +*) have other parameters, e.g., $\{p_i\}$, $\{\Sigma_i\}$ in GMD

Clusters can be represented by partition matrix (U) *or* prototypes $V = \{v_k\}$



$$U_c = \begin{matrix} 1 & 1 & \ldots & 1 & 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & 1 & 1 & \ldots & 1 \end{matrix}$$

Crisp : hard boundaries



$$U_f = \begin{matrix} .9 & 1 & \ldots & .6 & .3 & 0 & .2 & \ldots & 0 \\ .1 & 0 & \ldots & .4 & .7 & 1 & .8 & \ldots & 1 \end{matrix}$$

Fuzzy : soft boundaries

Rock Bass, Penghu, Taiwan

The c-Means Clustering Models

The c-Means Models : Early History

1802

Gauss uses least squares to estimate orbits of planets

Sends method to "a correspondent" in unpublished letter

1806

Legendre *publishes* first paper on least squares estimation

Gauss says *"been there, done that"*

Legendre *freaks out*

→ Gauss says Legendre *IS a freak!*

1812

Olbers produces Gauss's letter, it's date is authenticated

The c-Means Models : Modern History

1939	Tryon publishes first book on clustering
1948	Shannon develops basis of VQ = information theory
1956	Steinhaus discusses hard c-means functional for clustering
1957	Lloyd derives batch HCM/AO formulae algebraically
1962	Sebestyn formulates sequential hard c-means (SHCM)
1963	Ward gives a MSE "relational version" of batch HCM
1965	Zadeh publishes first paper on fuzzy sets
1967	MacQueen gives asymptotic statistical properties of SHCM
1969	Ruspini publishes first paper on fuzzy clustering
1973	Bezdek develops fuzzy c-means (FCM/AO) model/algorithm

Batch Hard and Fuzzy c-Means Models

Objective function

$$J_m(U,V) = \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}^m \|x_k - v_i\|_A^2$$

Inputs

Object data : $X=\{x_1,\ldots x_n\} \subset R^p$

Unknowns

(Fuzzy) Partition : $U \in M_{fcn}$

Prototypes $V=\{v_1,\ldots,v_c\} \subset R^{cp}$

Optimization Problem, m≥1

$$\underset{\substack{U\in M_{fcn}\\ V\in \Re^{cp}}}{\text{minimize}} \left\{ J_m(U,V) = \sum\sum u_{ik}^m \|x_k - v_i\|^2 \right\}$$

FONCs for extrema of the HCM/FCM Functionals

	FCM		HCM
Prototypes V=F(U,X)	$v_i = \dfrac{\sum_{k=1}^{n} (u_{ik})^m x_k}{\sum_{j=1}^{n} (u_{ij})^m}$	limit= $m \to 1^+$	$v_i = \dfrac{\sum_{k=1}^{n} u_{ik} x_k}{\sum_{k=1}^{n} u_{ik}}$
Partition U=G(V,X)	$u_{ik} = \left[\sum_{j=1}^{c} \left(d_{ikA} / d_{jkA} \right)^{\frac{2}{m-1}} \right]^{-1}$	limit= $m \to 1^+$	$u_{ik} = \begin{cases} 1 & d_{ikA} \le d_{jkA},\ j \ne i \\ 0 & \text{otherwise} \end{cases}$

$$d_{ikA} = \left\| x_k - v_i \right\|_A = \sqrt{(x_k - v_i)^T A (x_k - v_i)}$$

$A_{p \times p}$ positive definite

FONCs for extrema of J_1 provide a natural "equilibrium" between *U only* and *V only* models

U models: V=G(U)

$$v_i = \sum_{k=1}^{n} u_{ik} x_k \Big/ \sum_{k=1}^{n} u_{ik} = \sum_{x \in X_i} \frac{x}{n_i}$$

Centroid of cluster i

V models: U=F(V)

$$u_{ik} = \begin{cases} 1 & \|x_k - v_i\|_A \le \|x_k - v_j\|_A, \; j \ne i \\ 0 & \text{otherwise} \end{cases}$$

1-np classifier rule

HCM/FCM AO Algorithms

Input: *Unlabeled Object data*: $X \subset R^p$

User Picks: $c,\ m,\ \varepsilon,\ T,\ \|*\|_A,\ \|*\|_{err}$

Initialize:

$V_0 = (v_{10}, \ldots, v_{c0}) \in \Re^{cp}$

$U_0 = G(V_0, X)$

$V_1 = F(U_0, X)$ ----------> % For loop startup

$t=0$

AO Loop:

WHILE [$t<T$ and $\|V_{t+1} - V_t\|_{err} > \varepsilon$]

AO

$U_{t+1} = G(V_{t+1}, X)$ % Next partition

$V_{t+2} = F(U_{t+1}, X)$ % Next prototypes

WEND

Outputs: $(U^*, V^*) \in M_{fcn} \times \Re^{cp}$

HCM/FCM AO Algorithms

User choices (model parameters)

- inner product norm $\|*\|_A$ for J_m
- weighting exponent = $m \geq 1$
- c = # of clusters, $2 \leq c < n$

Most important choice you make !

Why? If you are in the wrong M_{fcn}, you *cannot* find the solution

User choices (control parameters)

- T = max. # of iterations
- ε = termination threshold
- termination norm $\|*\|_{err}$

HCM/FCM AO Algorithms

User choices
(iteration initialization)

(Fuzzy) Partition : $U_0 \in M_{fcn}$

or

Prototypes $V_0=\{v_{1,0},\ldots,v_{c,0}\} \subset R^{cp}$

Second most important choice you make !

Why? You are in th … rong solution !

True for *ANY* it … itial guess !

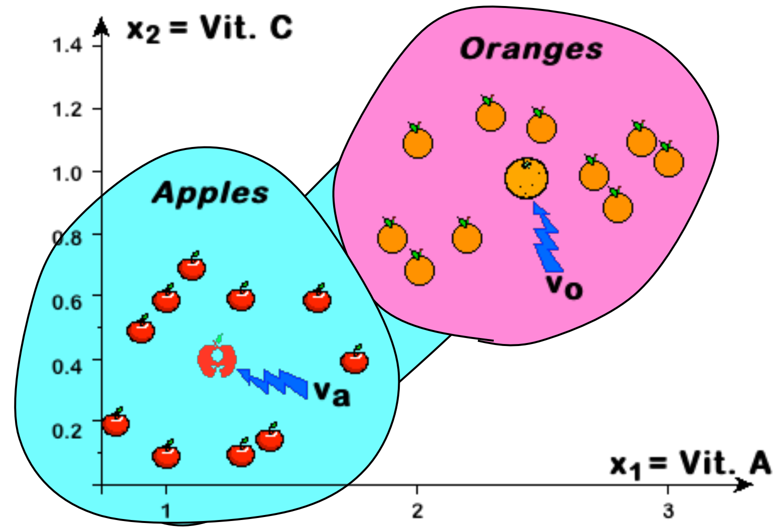
Hence, the … hms !

Initializing … han *U*

HCM/FCM Algorithms

X = 20 (unlabeled) points

x	x_1	x_2	Init U_{10}	Init U_{20}	HCM U_1	HCM U_2	FCM U_1	FCM U_2
1	1.00	0.60	1	0	1	0	0.97	0.03
2	1.75	0.40	1	0	1	0	0.77	0.23
3	1.30	0.10	1	0	1	0	0.96	0.04
4	0.80	0.20	1	0	1	0	0.94	0.06
5	1.10	0.70	1	0	1	0	0.95	0.05
6	1.30	0.60	1	0	1	0	0.97	0.03
7	0.90	0.50	1	0	1	0	0.96	0.04
8	1.60	0.60	1	0	1	0	0.84	0.16
9	1.40	0.15	1	0	1	0	0.95	0.05
10	1.00	0.10	1	0	1	0	0.95	0.05
11	2.00	0.70	1	0	0	1	0.33	0.67
12	2.00	1.10	1	0	0	1	0.19	0.81
13	1.90	0.80	1	0	0	1	0.39	0.59
14	2.20	0.80	1	0	0	1	0.10	0.90
15	2.30	1.20	1	0	0	1	0.04	0.96
16	2.50	1.15	1	0	0	1	0.01	0.99
17	2.70	1.00	0	1	0	1	0.01	0.99
18	2.90	1.10	0	1	0	1	0.05	0.95
19	2.80	0.90	0	1	0	1	0.03	0.97
20	3.00	1.05	0	1	0	1	0.06	0.94
	$\bar{V}_1$	$\bar{V}_2$	$v_{1,0}$	$v_{2,0}$	$v_{1,f}$	$v_{2,f}$	$v_{1,f}$	$v_{2,f}$
	1.21	2.43	1.57	2.85	1.21	2.43	1.21	2.50
	0.39	0.98	0.61	1.01	0.39	0.98	0.41	1.00

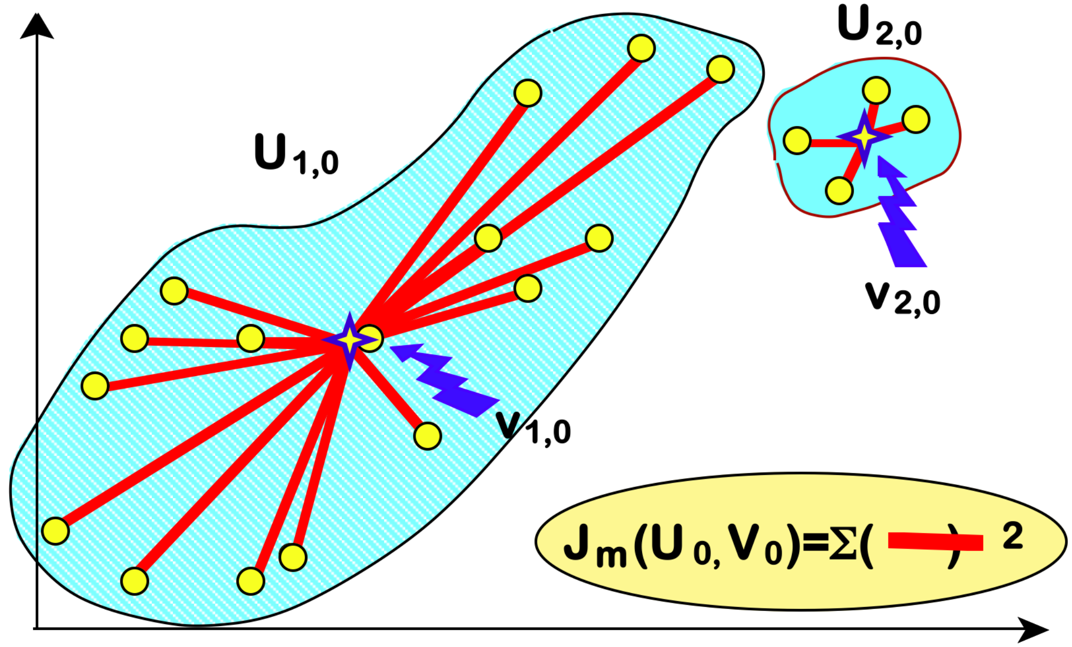


$H(U_{FCM}) = H_{MM} = U_{HCM}$

Not always true !

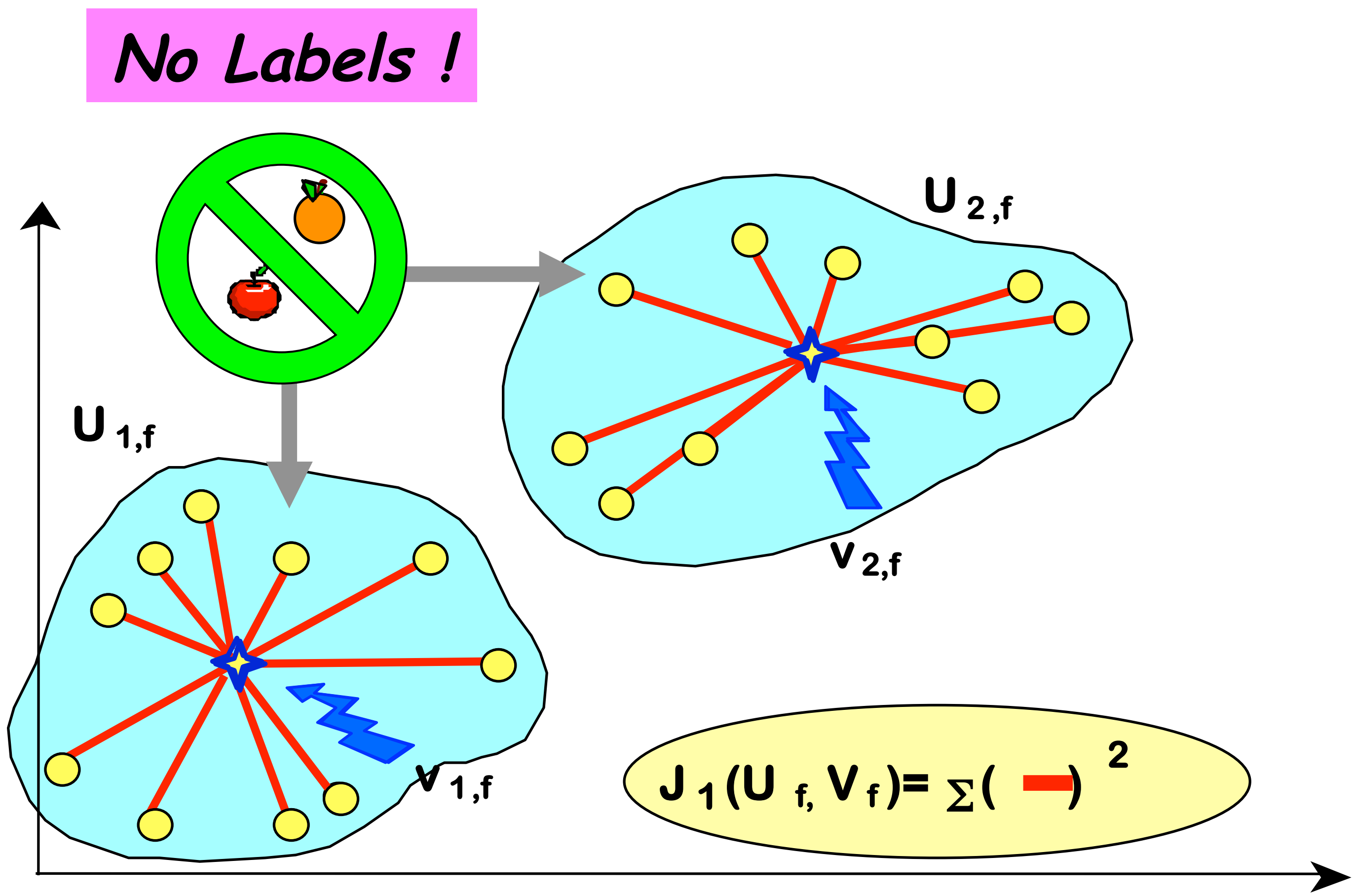
$V_{AVG} = V_{HCM} \sim V_{FCM}$

Geometry of the c-Means Models

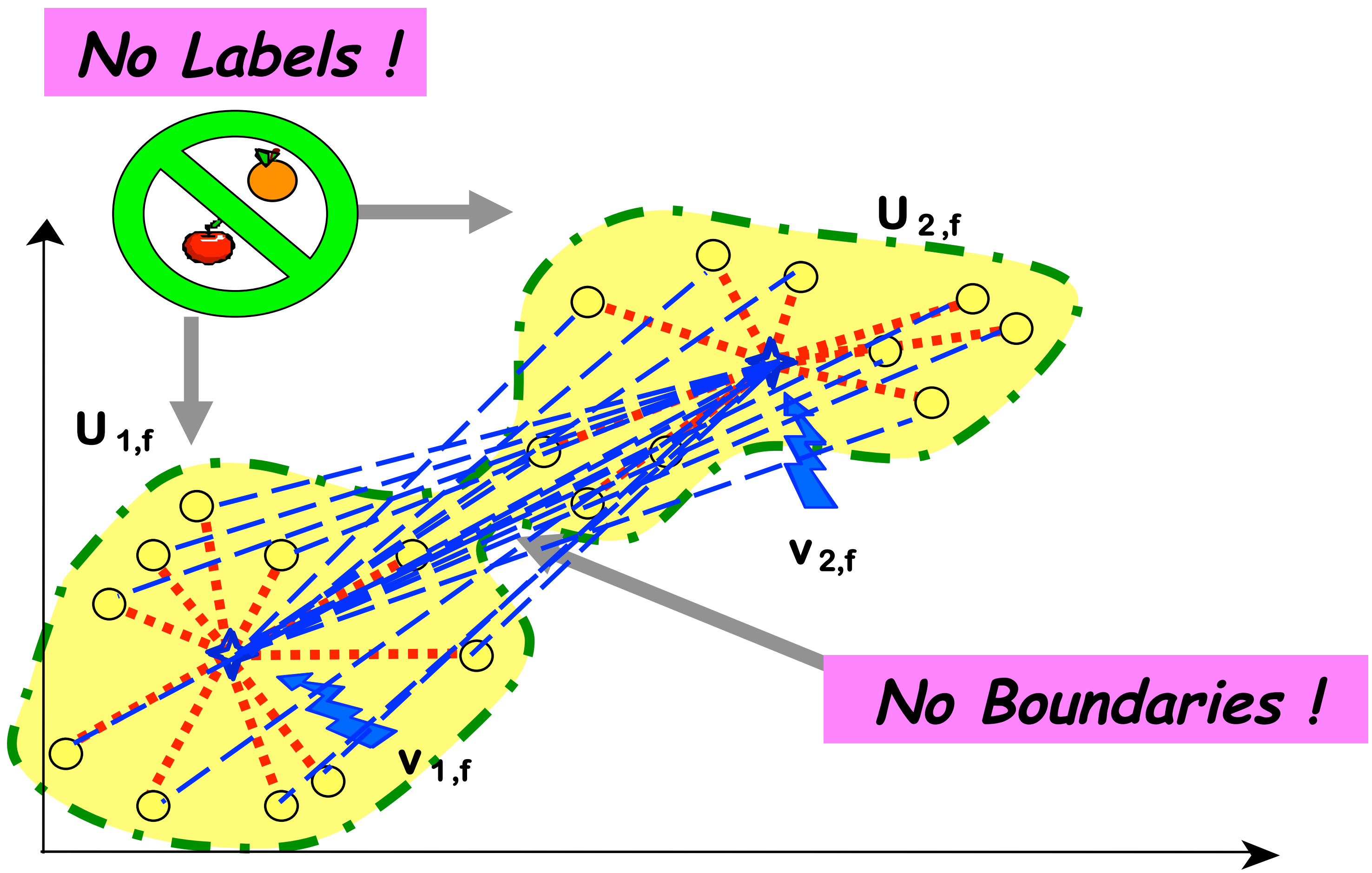


Initial Hard 2-partition for HCM and FCM

Terminal HCM 2-partition - 5 Iterations



Terminal FCM 2-partition - 6 Iterations



Clustering can be used for *Some* Medical Imaging Problems

Basic Image Processing

Segmentation

A. Unsupervised (Clustering)
B. Supervised (Classifiers)
C. Semi-Supervised (A $\Rightarrow$ B)

Medical Databases

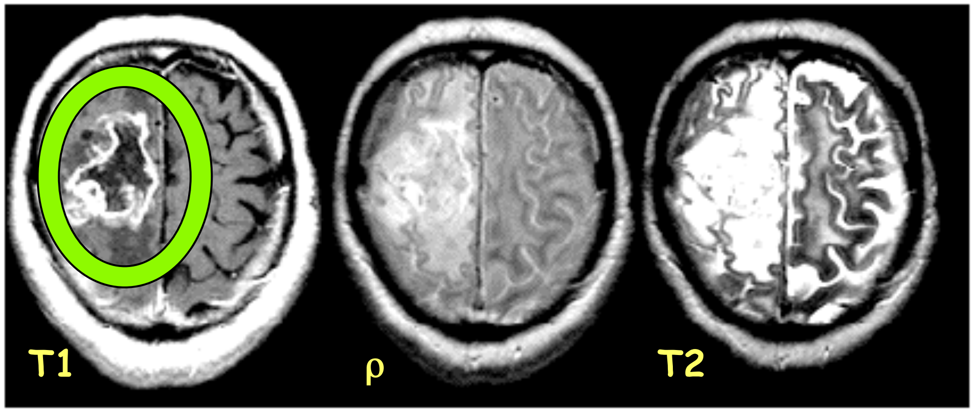
Query Retrieval
Patient Records
Image Matching

3-D diagnostic

Segmentation
Volume Estimation
Visualization
Surgical Planning
Robotic Surgery

Early MRI Segmentation

Bezdek, Hall, Clark, Goldgof, Clarke (1997). Medical image analysis with fuzzy models, *Stat. Meth. in Medical Research*, 6, 191-214.



$p=3 : x_{ij}=(T1_{ij}, \rho_{ij}, T2_{ij})$

One slice: n = 256x256

c = 7 for the k-nn rule

Radiologist GT (Ground Truth)

Supervised knn

Estimate of Tumor Pixels made by an optimized k-nn rule (retraining to best match GT)

Not very good

Early MRI Segmentation

Bezdek, Hall, Clark, Goldgof, Clarke (1997). Medical image analysis with fuzzy models, *Stat. Meth. in Medical Research*, 6, 191-214.

KB mask X* is made by segmenting X with FCM at c = 7

H(X) is a hardened segmentation of X

KB removes CSF, white matter, gray matter, skull, and air tissues from H(X), leaving X*

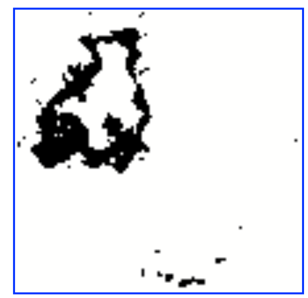
Finally, four methods are used to segment X* at c = 5

Radiologist GT (Ground Truth)

FCM

FCM + VGC

Unsupervised



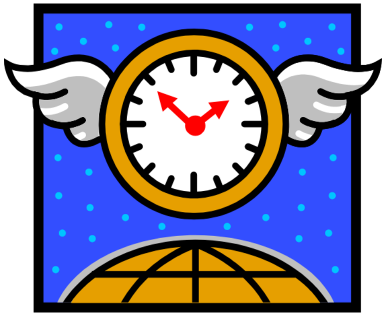
ssFCM

KB

Supervised. Training data chosen by the KB system

2 important issues when using c-means/AO and/or GMD/AO for image segmentation:

Speed



Scalability

Sockeye Salmon, Seattle, WA

These methods are given for FCM, …. but most of them will also work with HCM, PCM, ssFCM, and GMD/EM

Acceleration of c-Means

(Selected) Acceleration tricks for FCM when X_L is *Loadable*

Change Algorithm

1986	Cannon/Dave/Bezdek propose *approximate* FCM (AFCM)

Change Implementation

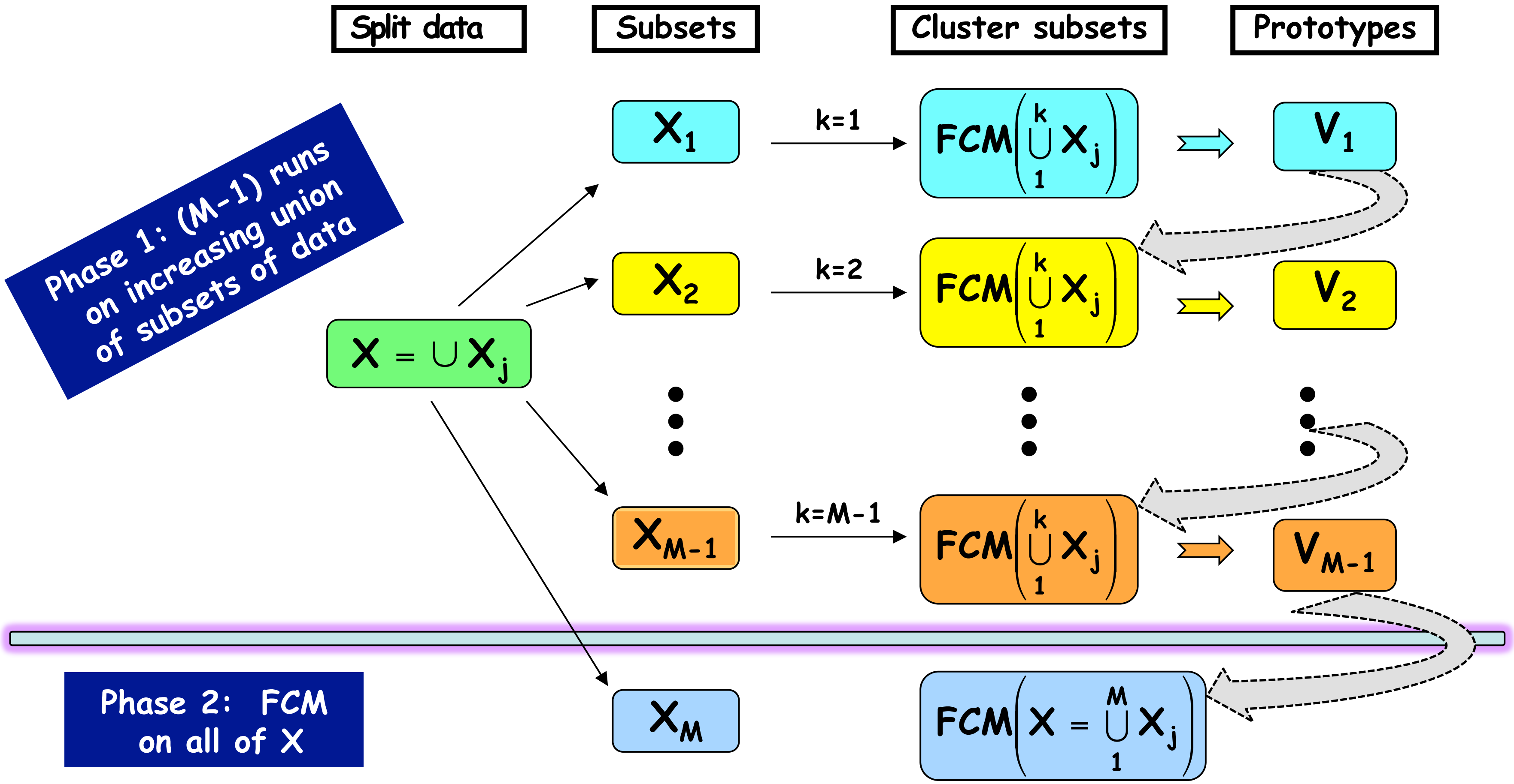
1998	Cheng/Goldgof/Hall define multistage init. for FCM (mrFCM)
1994	Kamel/Selim incrementally update (U. V), hidden U implicit
2002	Kolen/Hutcheson study hidden U explicitly and extensively
1999	Altman gives one U update trick, and one with "mini" mrFCM
2003	Borgelt and Kruse accelerate FCM with five NN tricks
2008	Anderson, Luke and Keller implement FCM on a GPU- *fast !*

Nutshell Comparison of Acceleration Tricks For *Loadable* Data

Method	p	c	n	Data sets	Ave. Speedup
Cannon/DB-AFCM	10	10	0.25 mb	8 bit Landsat	6:1
Kamel/S – ΔU, ΔV	Varies	Varies	Small	British Towns German Towns Fossil, Wine	1.2 : 1
Cheng/GH-mrFCM	3 6	10 10	0.25 mb 0.55 mb	8 bit MRI 8 bit Landsat	3:1 3:1
Altman 1- ΔU Altman 2-mr1FCM	3 3	3 3	1 mb 1 mb	8 bit Landsat 8 bit Landsat	3:1 10:1
Kolen/H – No U	9	10	20 mb	8 bit hypoth.	9:1
Borgelt/K NN tricks 3: Step Width Exp. 4: Momentum 5: Self Adaptive LR	8 9 4 13	3 2 3 3	4177 286 150 178	Abalone Breast Canc Iris Wine	2:1 2:1 2:1 2:1
Anderson/LK-GPU	4-32	4-64	64-8192	Rnd Spherical with Rnd means	10-100:1
Eschrich/KHG-brFCM	3 2	10 5	0.25 mb 0.4 mb	12 bit MRIs 8 bit IR images	290:1 60-306:1

mrFCM="multi stage" FCM

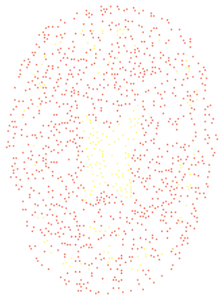
Cheng, T. W., Goldgof, D. B., & Hall, L. O. (1998). Fast fuzzy clustering. *Fuzzy Sets and Systems, 93, 49-56.*



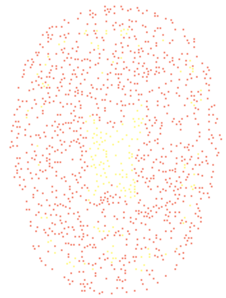
Altman's Method: M = 2 (1 stage of preprocessing)

MRI segmented by mrFCM

Cheng, T. W., Goldgof, D. B., & Hall, L. O. (*1998*). Fast fuzzy clustering. *Fuzzy Sets and Systems, 93, 49-56.*



Stage 1



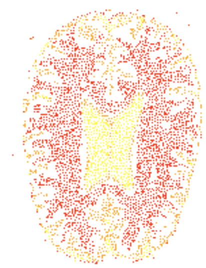
Stage 2

Stage 3

Stage 4

Stage 5

Stage 6

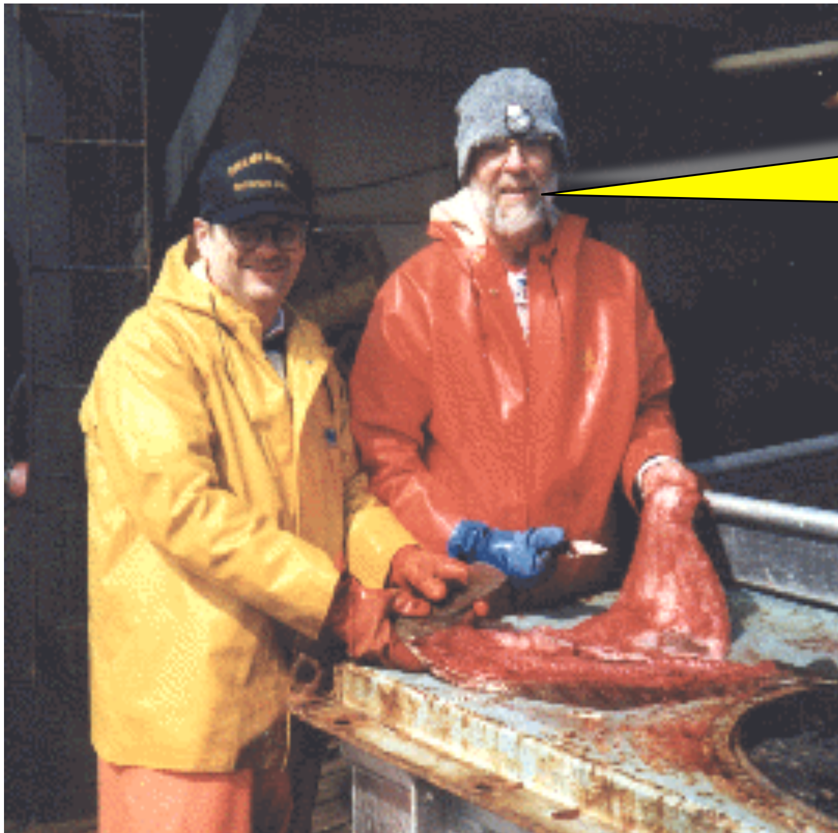


Stage 7

Last Pass Full FCM

Acceleration is due to *improved initialization* of FCM by using final V from stage k as initial V for stage k+1

Speedup ~ 3:1



Using FCM/EM with VL Data

Ling Cod, Alaska, USA

Very Large (VL) Data

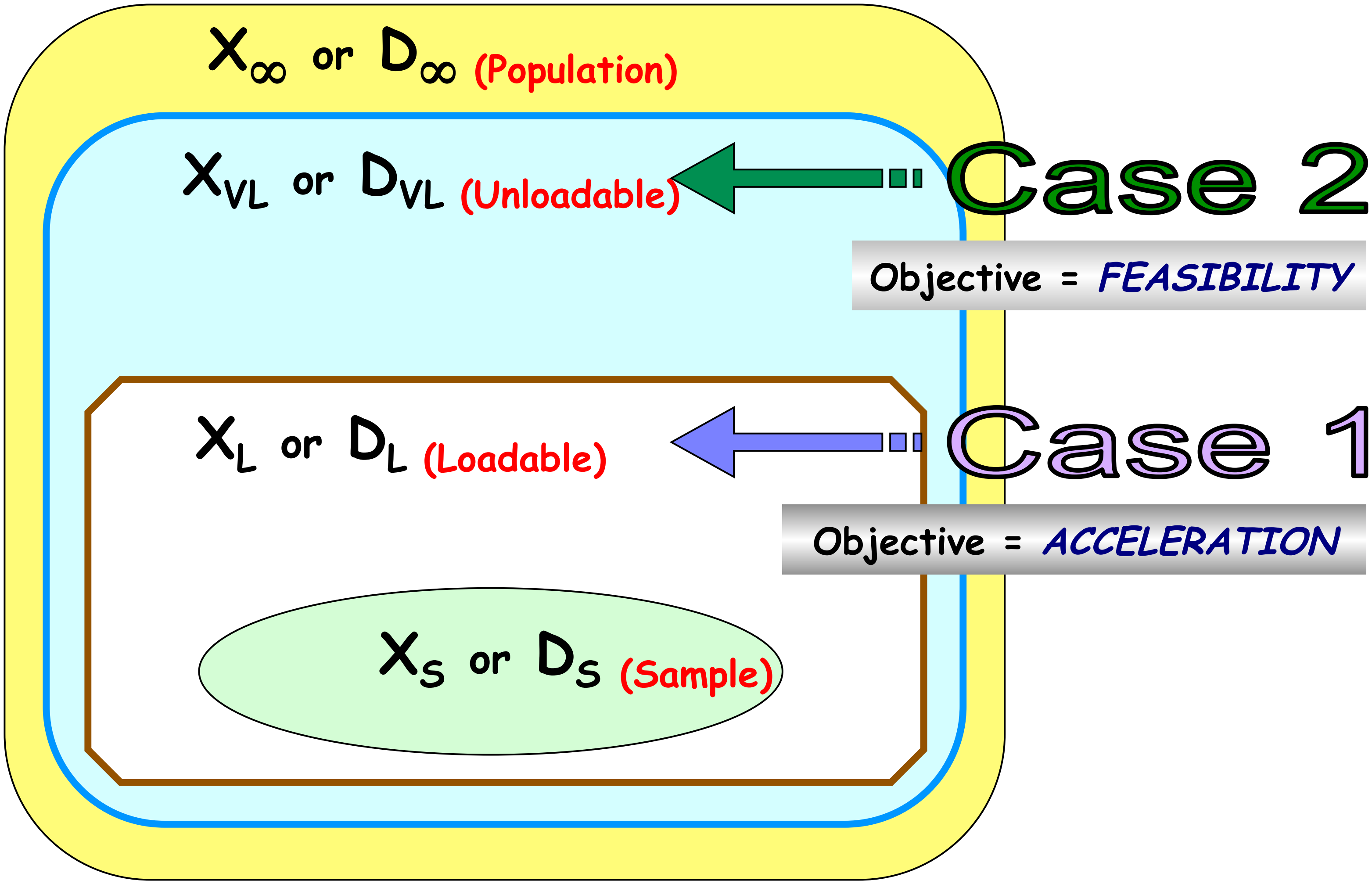
How Big is (VL) Data ?

Bytes	"Data Size (1996,Huber)"
10^2	Tiny
10^4	Small
10^6	Mega - medium
10^9	Giga - large
10^{10}	Huge
10^{12}	Tera - Monster
$10^{>12}$	Very Large (VL)

We can't cluster data this big (in a single computer) : VL is "unloadable"… so …

Most VL methods build "cluster-friendly" (loadable) subsets by *sampling* or *splitting*

2 Sampling sub-cases



Case 1

If X_L or D_L is *Loadable* we can *Quantitatively* *Compare* Lit-Clusters ↔ Approx. Clusters

Case 2

If X_{VL} or D_{VL} is *Unloadable* *Comparison is impossible*

so ...

case 2 validity rests with "good" case 1 examples

Case 2 VL methods may provide acceleration *and* feasibility

Case 2 If X_{VL} or D_{VL} is *Unloadable* …

there are 2 basic approaches to scaling up

1. Progressive (or) Random sampling + non-iterative extension + completion

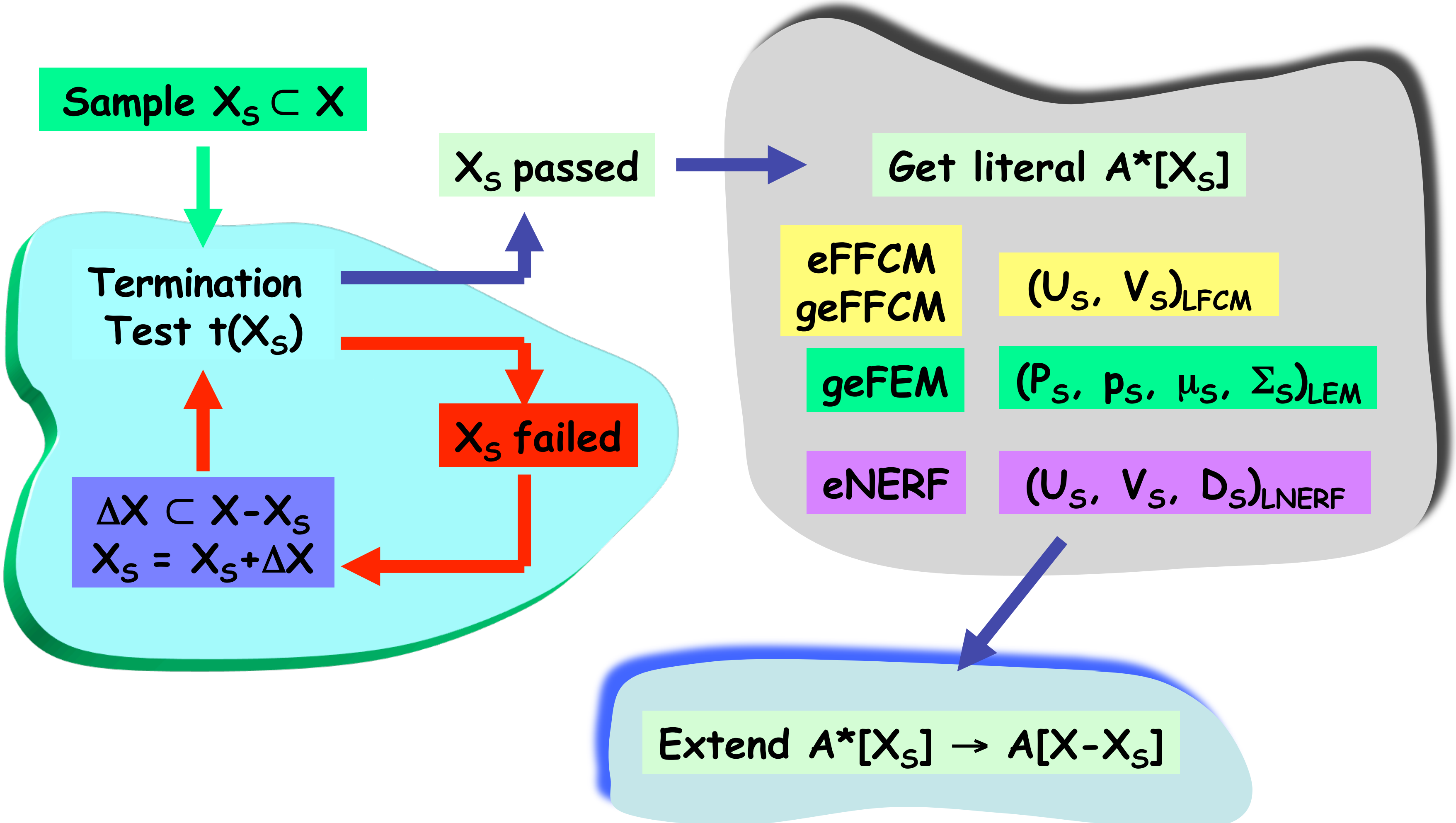
2. Incremental (aka "ensemble") clustering using chunks of distributed data

Many of these methods can be used with other pattern recognition algorithms. For clustering, we let …

A*= [U*, V*] = exact (literal) partition and prototypes

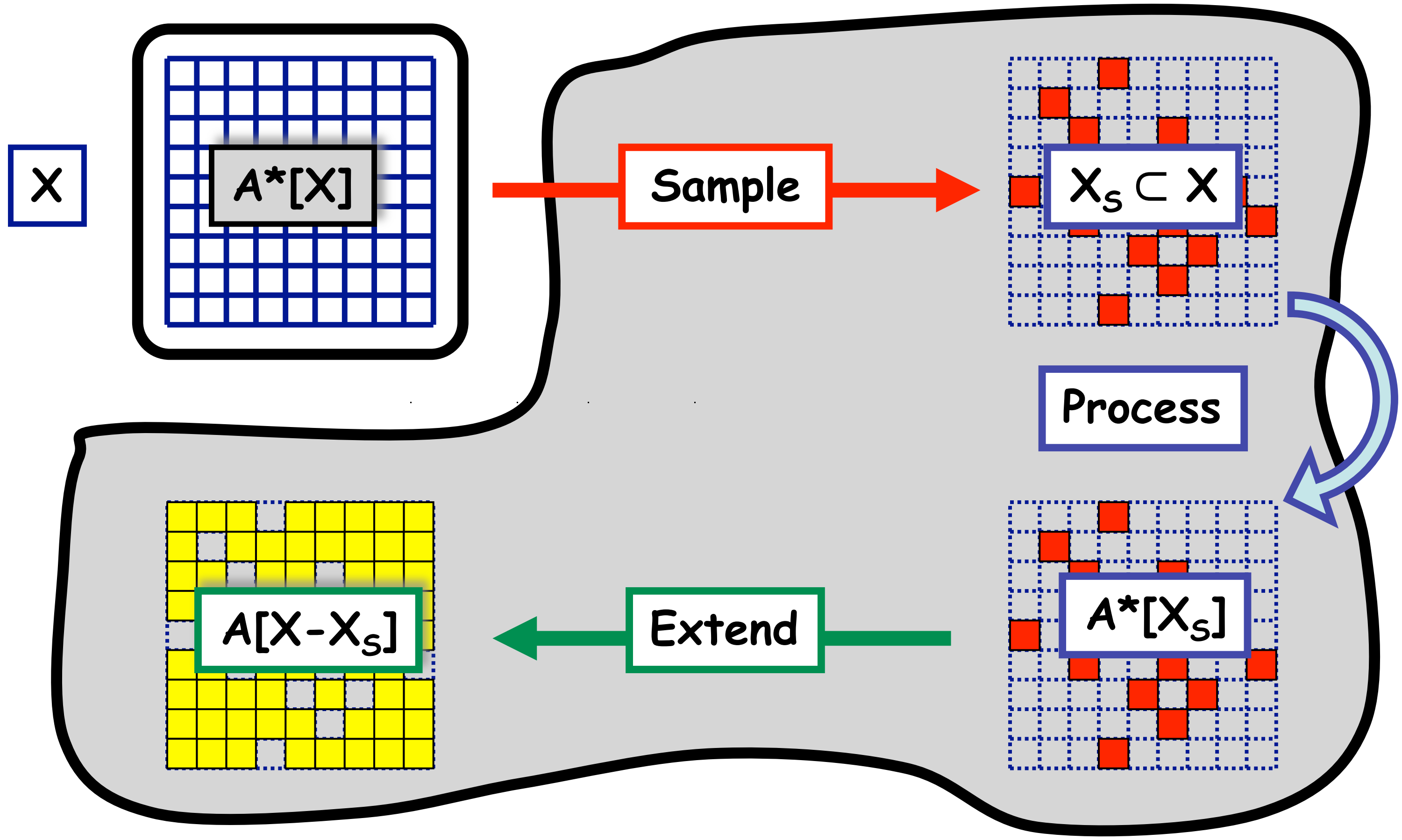
A = [U, V] = (any) approximation to A* by (1) or (2)

What is *Progressive Sampling* ?



What is *Extension* ?

Algorithm $A : X \subset \Re^p \mapsto A[X] \subset \Re^q$



Literal A*[X] $\cong$ Approx. $A(X)=A^*[X_S] \| A[X-X_S]$

Sample $X_S \subset X$

Process $FCM[X_S]$

Extend $FCM[X_S] \rightarrow FCM[X-X_S]$

(Non-Iterative) Generalized extension of *Fuzzy c-Means* [FCM → eFFCM/geFFCM]

with prototypes V_S and $x_k \in X-X_S$

$(X - X_S) \ni x_k$

$\{\cdots v_{i,S} \cdots v_{j,S} \cdots\} = V_S$

$$u_{ik} = \left[\sum_{j=1}^{c} \left(\|x_k - v_{i,S}\|_A \Big/ \|x_k - v_{j,S}\|_A \right)^{\frac{2}{m-1}} \right]^{-1} = \varphi(V_S, X - X_S)$$

FONC for U to min J_m

Is a *classifier* on X-X_s, trained *w/o labels* on X_s !

5 Progressive Sampling+Extension (pse) Schemes for FCM/EM

2002, eFFCM

Pal/ Bezdek: …plexity reduction for "large image" proce… …EE SMC*, B-32(5).

2006, geFFCM & geFE…

Hathaway/Bezdek: Fuzzy and pro… …clustering in VL data sets, *Comp. Stat. And Ana…*

2006, eNERF

Bezdek/Hathaway/Huband/Leckie/Rao: Approximate clustering in VL relational data, *Int. JIS*, 21.

2008, SS1

Wang/Bezdek/Leckie/Kotagiri: Selective sampling for approx clustering of VL data, Int. JIS, 23(3).

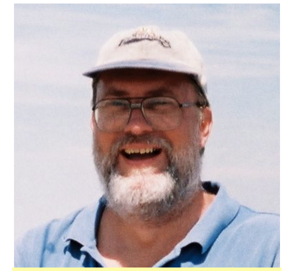
2011, SS2

Wang/Leckie/Kotagiri/Bezdek: Approx pairwise clustering for large data sets via sampling plus extension, Patt. Recog., 44.

For large n, random sampling+extension (rse) may be better !

Nik Pal

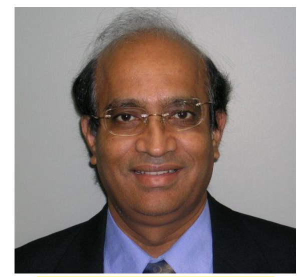
Jim Bezdek



Rick Hathaway

Jackie Huband

C… …kie



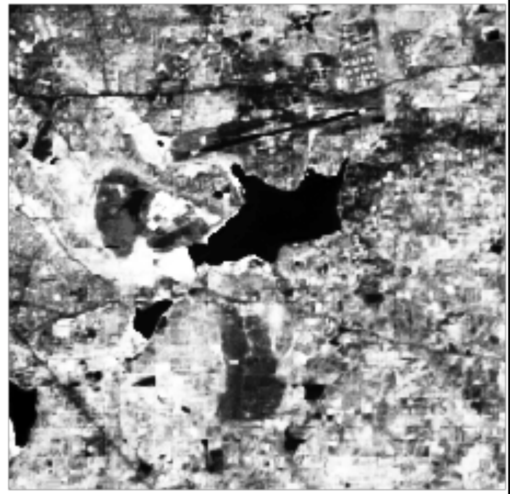
Rao Kotagiri

Liang Wang

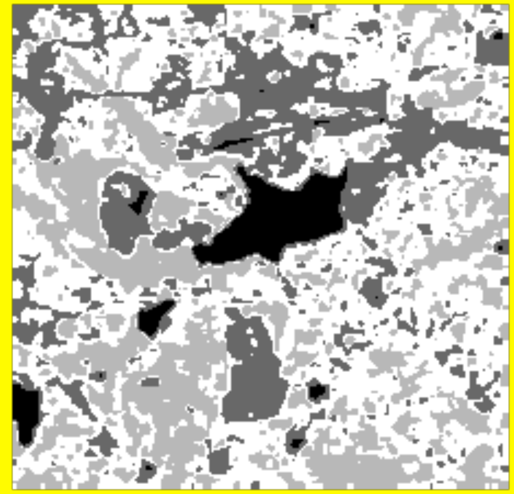
eFFCM on a Loadable Image

Indian Satellite (Landsat Image)

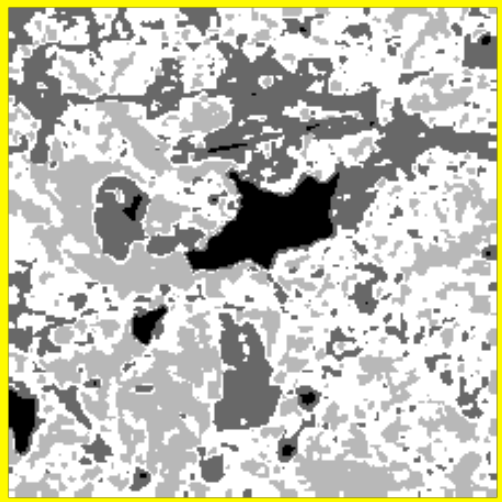
COMPARE 9% vs 100% !



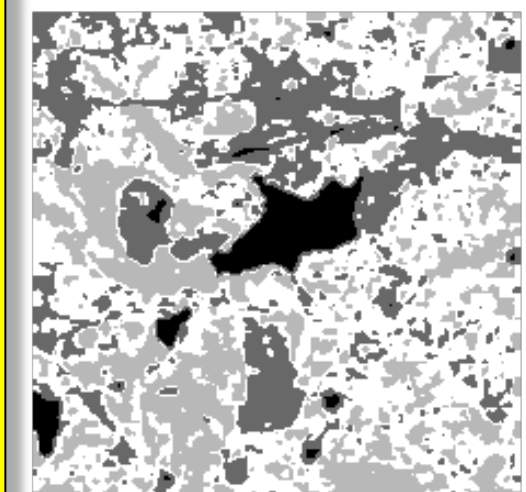
Input image
256x256x256



LFCM 100% of data
mrFCM 100% of data

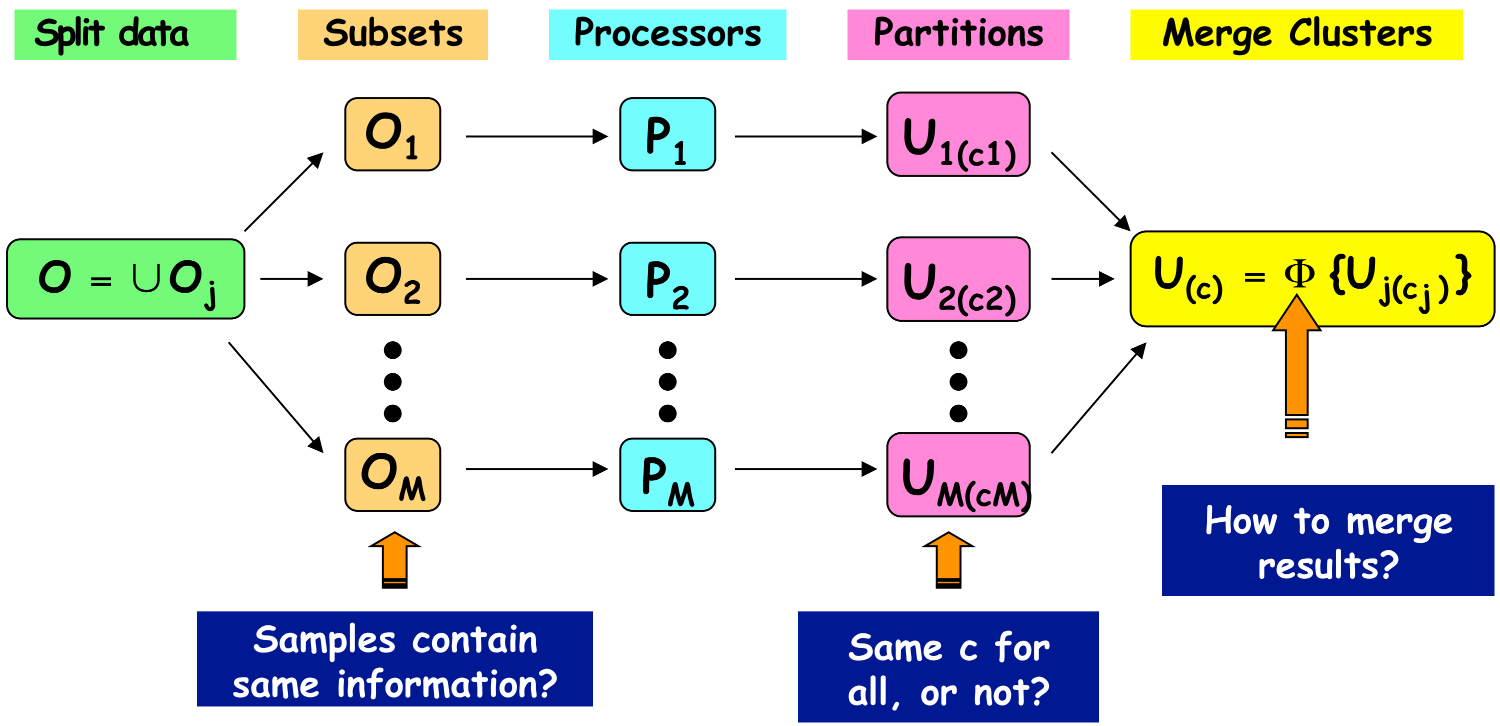


eFFCM
9% of data, div only



eFFCM
29% of data, div & $\chi 2$

Incremental/Distributed Clustering in VL Data



Problems for the Distributed Clustering Approach

Incremental/Distributed Clustering in VL Data

brFCM = "bit reduct." FCM

Eschrich/Ke/Hall/Goldgof (*2003, brFCM*). Fast accurate fuzzy clustering through data reduction. *IEEE TFS*, 11(2), 262–270.

Compression *for image data*, uses wFCM algorithm, (loadable) implementation

spFCM = "single pass" FCM

Hore, Hall, Goldgof (*2007, spFCM*). Single pass fuzzy c- means, Proc. FUZZ-IEEE 2007, 1-7.

Uses wFCM algorithm, partially distributed VL implementation

oFCM = "on line" FCM

Hore/Hall/Goldgof/Gu/Maudsley/Darkazanli (*2009, oFCM*). A scalable framework for segmenting MRIs, *J Signal Proc. Syst*. 54(1-3), 183–203.

Uses wFCM algorithm, fully distributed VL implementation

Weighted FCM = wFCM

wFCM

$$\min_{(U,V)}\left\{J_{mw}(U,V:X)=\sum_{k=1}^{n}\sum_{i=1}^{c}w_k(u_{ik})^m\left\|x_k-v_i\right\|_A^2\right\}$$

Inputs

$X\subset R^p$ + n fixed weights $\{w_k\}\subset(0,\infty)$

Unknowns

$U\in M_{fcn}$ + $V=\{v_1,\ldots,v_c\}\subset R^{cp}$

Partition $U=G(V,X)$

$$u_{ik}=\left[\sum_{j=1}^{c}\left(d_{ikA}/d_{jkA}\right)^{\frac{2}{m-1}}\right]^{-1}$$

FONCs

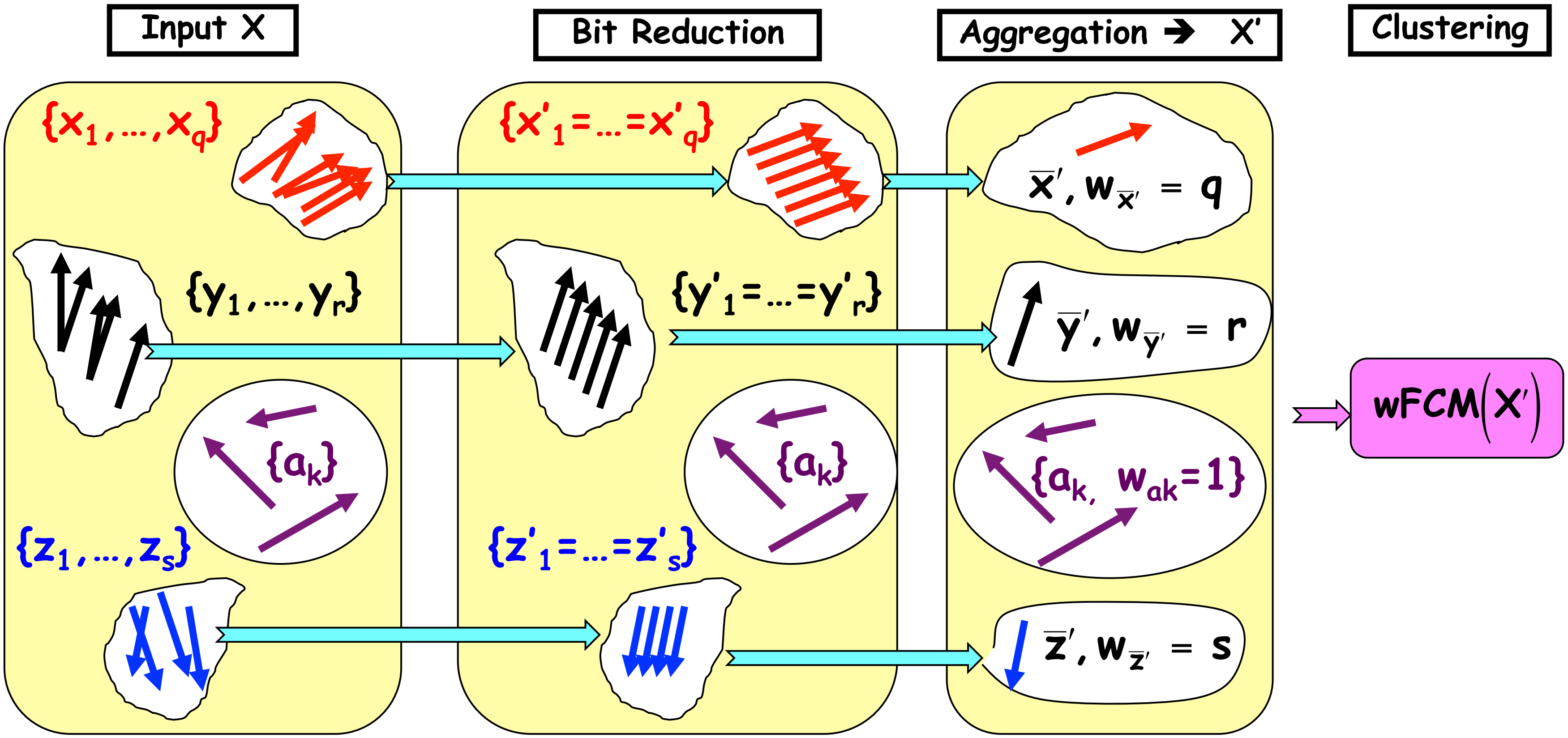
Prototypes $V=F(U,X)$

$$v_i=\frac{\sum_{k=1}^{n}w_i(u_{ik})^m x_k}{\sum_{j=1}^{n}w_i(u_{ij})^m}$$

ONLY Change (Bezdek, 1981)

brFCM="bit reduction" FCM

Eschrich/Ke/Hall/Goldgof (2003). Fast accurate fuzzy clustering through data reduction. *IEEE TFS*, 11(2), 262-270.



Not *explicitly* for VL data, X must be fully loadable, but X' may be loadable even if X is unloadable !

spFCM="single pass" and oFCM = "on line" FCM

Split data

$$X = \cup X_j \quad : n = \sum n_j$$

First pass

$$FCM(X_1) = (U, V)$$

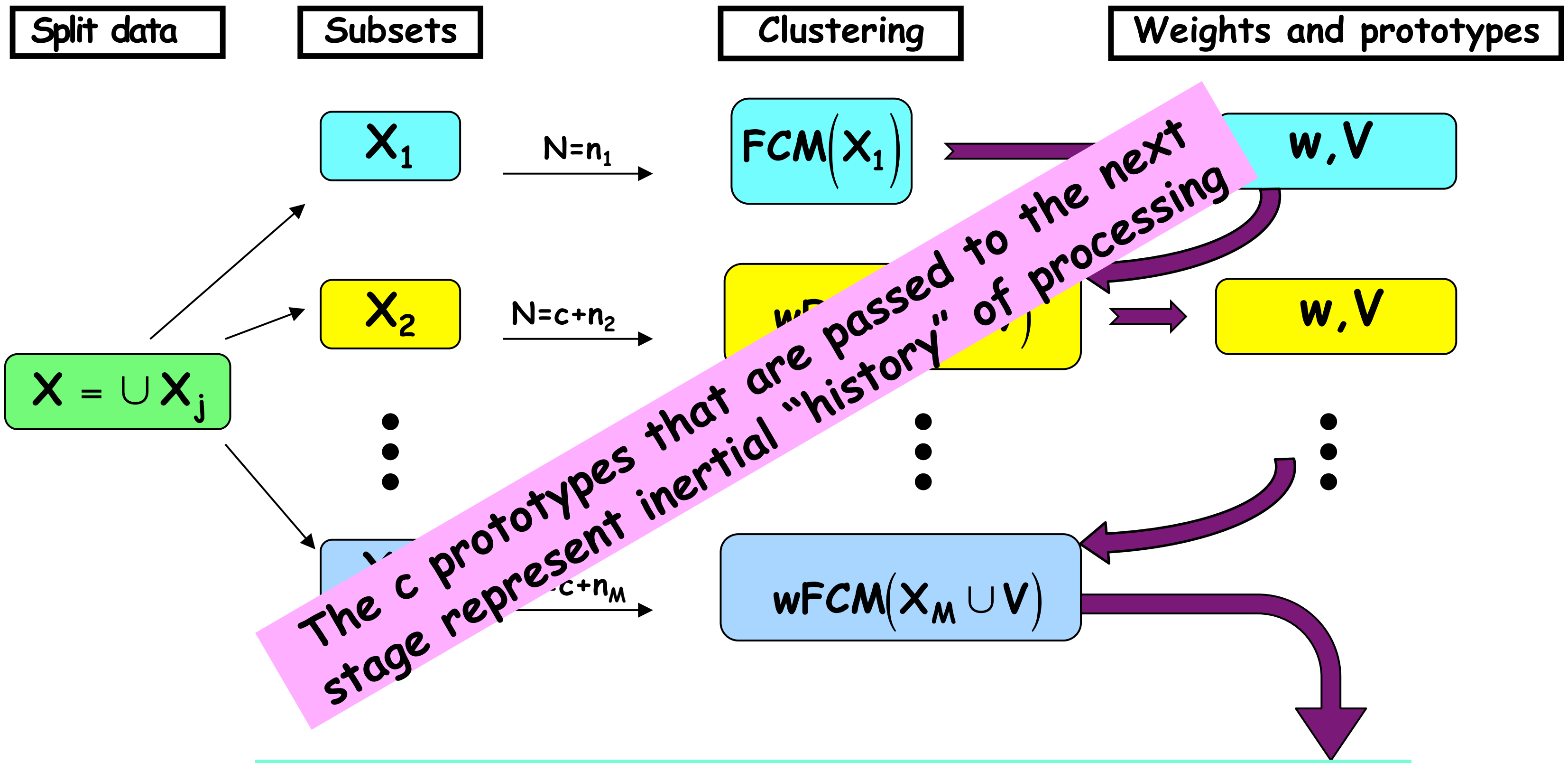
Rowsums of U
after **pass j ≥1**

$$\omega_i = \sum_{k=1}^{n_j} u_{ik} ; 1 \le i \le c$$

weights for wFCM
before **pass j>1**

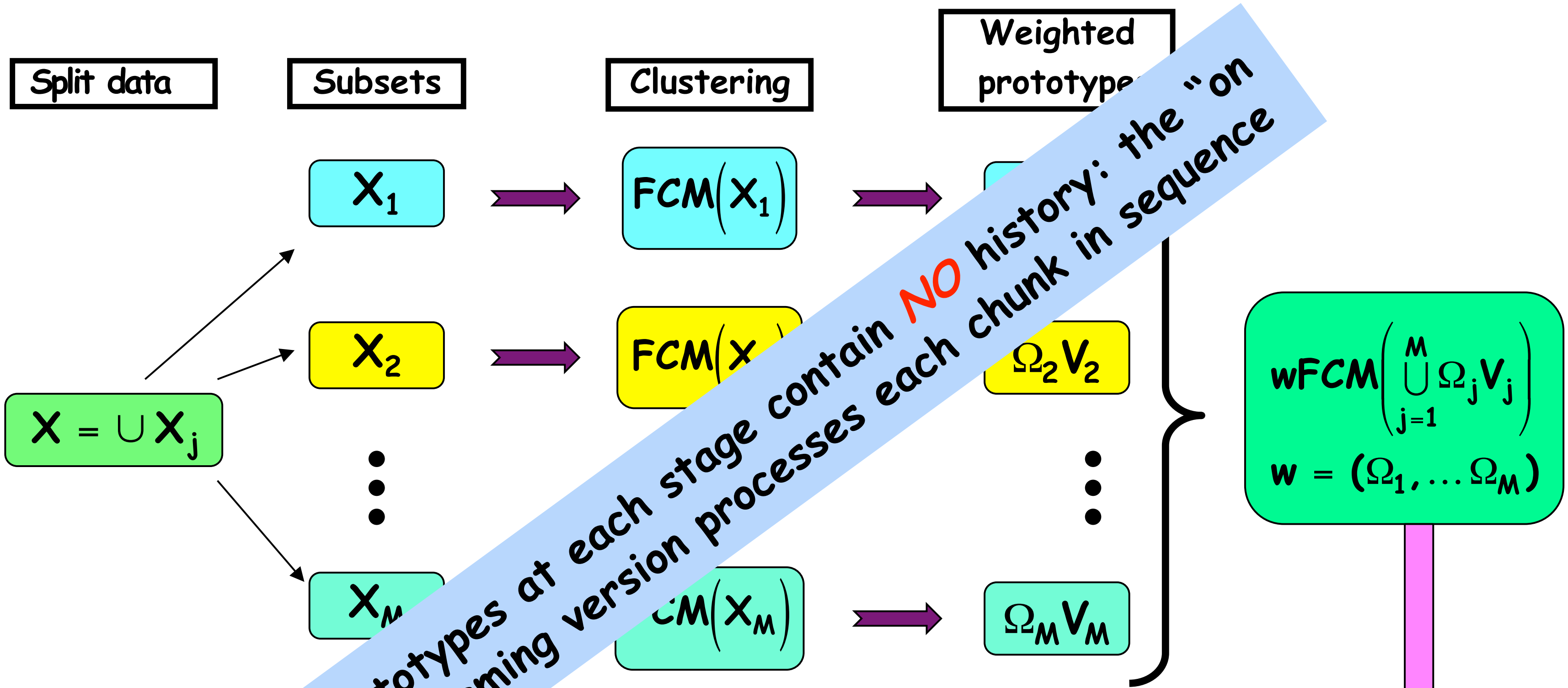
$$w = ([1], \Omega) = (\underbrace{1, 1, \ldots, 1}_{n_j \text{ times}}, \omega_1, \ldots, \omega_c)$$

Architecture of spFCM : c is chosen and fixed by user



A partition of X_{VL} is built using the HCM/FCM FONCs to compute U with the final V's

Architecture of oFCM : c = "max" is same for all blocks

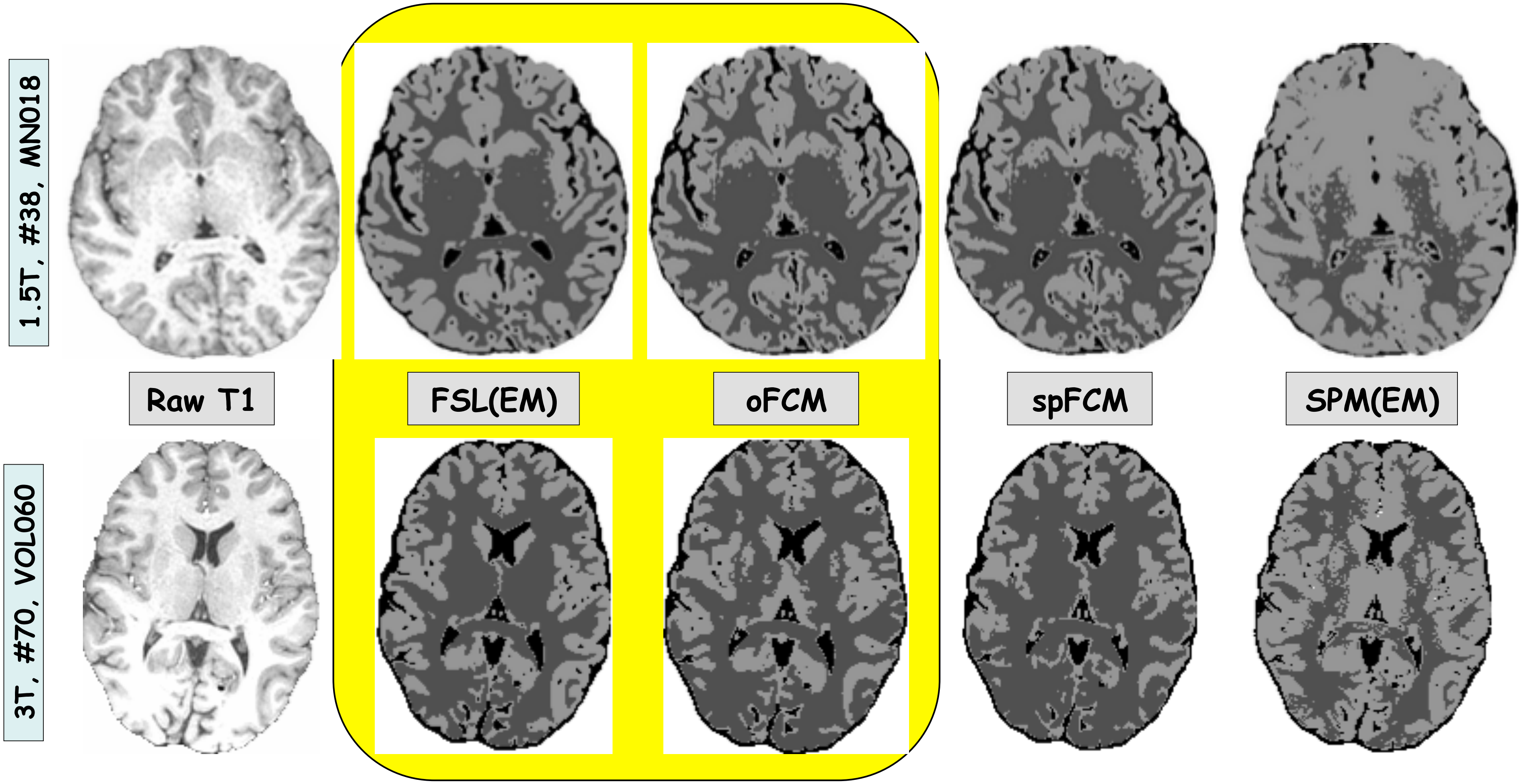


The c prototypes at each stage contain **NO** history: the "on line" or streaming version processes each chunk in sequence

… fuzzy partition of X_{VL} is built using the …CM/FCM FONCs for U with the final V's

Visual comparison of segmentations by 4 algorithms

FSL is GMD/EM-AO



FSL ~ oFCM best segment raw images : SPM worst

Comparing rseFCM, spFCM & oFCM to LFCM

Havens/Bezdek/Hall/Leckie/Palaniswami (2011).
Fuzzy c-Means algorithms for very large data, (in review TKDE).

Havens/Palani/Bezdek | Hall | Leckie

3 Evaluation Criteria

Run time (time LFCM/time VL Approximation)	9 data sets
Adjusted Rand Index ($0 \leq ARI \leq 1$): *2 cases* 1. matches hardened H(U) to GT partition V 2. (implicitly) matches H(U) to $H(U_{LFCM})$	 Labeled data Unlabeled data
Fuzzy ARI ($0 \leq fARI < 1$) matches approximate fuzzy U to fuzzy U_{LFCM}	6 unlabeled MRI image data sets

All outputs are averages of 21 random samplings/initializations

MRI Image Data n ~ 4×10^6, p=1 or 3, c=3

p=1	0.1% samples SU	0.1% samples ARI*	0.1% samples fARI	1% samples SU	1% samples ARI*	1% samples fARI	10% samples SU	10% samples ARI*	10% samples fARI
rseFCM	22	0.97	0.66	18	0.99	0.66	7	1	0.66
spFCM	13	0.98	0.66	13	0.98	0.66	8	0.98	0.66
oFCM	2	1	0.66	4	1	0.66	4	1	0.66
brFCM	108	1	0.66	50	1	0.66	8	1	0.66

p=3	0.1% samples SU	0.1% samples ARI*	0.1% samples fARI	1% samples SU	1% samples ARI*	1% samples fARI	10% samples SU	10% samples ARI*	10% samples fARI
rseFCM	29	0.97	0.47	24	1	0.47	8	1	0.47
spFCM	18	0.96	0.46	13	0.96	0.46	7	0.96	0.46
oFCM	2	0.78	0.38	2	0.93	0.44	3	1	0.47

ARI* matches H(U) to $H(U_{LFCM})$

Empirical Conclusions: pseFCM & rseFCM

Sampling → Two types (random and progressive). Easily adaptable for extensions to VL data with *many* other algorithms

Extension → Non-iterative scaling for *many* algorithms

rseFCM →
- Superiority to pseFCM increases with n
- Faster than spFCM/oFCM for large n
- Average speedup of LFCM ~ 30:1
- Good Approximation to LFCM clusters

Empirical Conclusions: brFCM, spFCM & oFCM

brFCM → Excellent acceleration for many images

Average speedup of brFCM ~ 100:1

spFCM → Retains "history" of clusters as more data chunks are added to processing

oFCM → No history retention; useful for on-line streaming analysis (of chunks)

Two important topics I did not discuss

1. Cluster Validity

The hardest and most important question !

The "right" value of c in truly unlabeled data?

Easy to see there is no general solution !

2. Clustering in VL (unloadable) data

∃ ≈ solutions for object & relational data

Needed especially in VL data mining

Distributed vs. Progressive Sampling

Este soy yo !

1980

1984

Oye papá, envolverlo

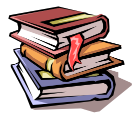
1985

1990

¡Despiértese!

¡Es terminado!

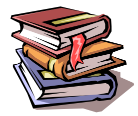

Ya esta, chicos !
(That's it, folks)
Muchas
Gracias !
Adios!



(U, V) Models

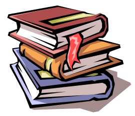
Min VQ Error	Shannon, C. E. (1948). A mathematical theory of communication, *Bell Systems Tech. Jo.*, 27, 623-656.
HCM-AO begins	Lloyd, S.P. (1957). Least squares quantization of PCM, [originally an unpublished Bell Labs technical note], reprinted in *IEEE Trans. IT*, 28, March, 1982], 129-137.
ISODATA + HCM	Ball, G. and Hall, D. A. (1967). A clustering technique for summarizing multivariate data, *Behav. Sci.*, 12, 153-155.
FCM (m=2)	Dunn, J.C. (1974). A Fuzzy Relative of the ISODATA Process and its use in Detecting Compact, Well-Separated Clusters, *Jo. Cybernetics*, 3, 1974, 32-57.
FCM (m>1)	Bezdek, J. C. (1973). Pattern Classification with Fuzzy Sets, PhD Thesis, Cornell, Ithaca, NY

VQ**



(U, V) Models

GK	Gustafson, D. and Kessel, W. Fuzzy Clustering with a Fuzzy Covariance Matrix, in *Proc. IEEE CDC*, 1978, 761-766.
AO conv begins.	Bezdek, J.C. A Convergence Theorem for the Fuzzy ISODATA Clustering Algorithms, *IEEE Trans. PAMI.*, PAMI-2(1), 1-8. 1980.
FCL FCV	Bezdek, J.C., Coray, C., Gunderson, R. and Watson, J. Detection and Characterization of Cluster Substructure, I and II, *SIAM Jo. of Appl. Math.*, 40(2), 1981, 339-372.
AP	Windham, M.P. Geometric Fuzzy Clustering Algorithms, Fuzzy Sets and Systems, 3, 271-280, 1983.
ssFCM	Pedrycz, W. (1985). Algorithms of fuzzy clustering with partial supervision, *Pattern Recognition Letters*, 3, 13-20.
VQ	Gersho, A. and Gray, R. (1992). *Vector Quantization and Signal Compression*, Kluwer, Boston.
FC shells	Dave, R. and Bhaswan, K. Adaptive Fuzzy c-Shells Clustering and Detection of Ellipses, *IEEE Tran. Neural Networks*, 3(5), 1992, 643-662.
PCM	Krishnapuram, R. and Keller, J. A Possibilistic Approach to Clustering, *IEEE Trans. Fuzzy Systems*, 1(2), 1993, 98-110.
FCRM	Hathaway, R. and Bezdek, J.C. Switching Regression Models and Fuzzy Clustering, *IEEE Trans. Fuzzy Systems*, 1(3), 1993, 195-204.
NISP	Bezdek, J.C., Hathaway, R. and Pal, N.R. Norm-Induced Shell-Prototype (NISP) Clustering Models, *IEEE Trans. Fuzzy Systems*, 1995.
VQ	Gray, R. M. and Neuhoff, D. L. (1998). Quantization, IEEE Trans. IT, 44 (6), 2325-2383.
AO Conv.	Bezdek, J. C. and R. J. Hathaway (2003). Convergence of alternating optimization, Neural, Parallel and Scientific Computations, 11, 351-368.



(U, V, +) Models

Mixture of 2 1-D Normals	Pearson, K. (1894). Contributions to the mathematical theory of evolution, *Phil. Trans. Royal Soc. of London*, 185, 71-110.
GMD (any p)	Wolfe, J. H. (1970). Pattern clustering by multivariate mixture analysis, *Multivariate Behavioral Research*, 5, 329-350.
(Best) EM Survey	Redner, R.A. and H.F. Walker (1984). Mixture densities, maximum-likelihood, and the EM algorithm, *SIAM Review*, 26, 195-239.
EM Theory	Hathaway, R. J., Redner, R. and J. C. Bezdek (1987). Estimating the Parameters of Mixture Models with Modal Estimators, *Comm. in Stat. (A)*, 16(9), 1987, 2639-2660.
Mixtures Text	Titterington, D., Smith, A. and U. Makov (1985). <u>Statistical Analysis of Finite Mixture Distributions</u>, Wiley, NY.
Mixtures Text	McLachlan, G. and D. Peel (2000). <u>Finite Mixture Models</u>, John Wiley & Sons, New York, NY.