

# Bases de données multimédia

Recherche avec contrôle de pertinence  
Recherche pluri-modale  
Catégorisation de bases d'images

Michel Crucianu (CNAM)

<http://cedric.cnam.fr/~crucianm/bdm.html>

## Contenu de la séance

- Recherche itérative avec contrôle de pertinence
  - ◆ Critères de sélection et apprentissage actif
  - ◆ Évaluation du contrôle de pertinence et exemples
  - ◆ Structures d'index pour la recherche itérative
- Recherche pluri-modale
  - ◆ Descripteurs issus des données textuelles
  - ◆ Extraction et exploitation de relations statistiques
  - ◆ Extraction et exploitation de relations conceptuelles
- Catégorisation de bases d'images
  - ◆ Classification automatique
  - ◆ Méthodes semi-supervisées

## Contrôle de pertinence : motivation

- **Fossé sémantique** (*semantic gap*) entre
  - ◆ descripteurs du contenu visuel extraits automatiquement des images et
  - ◆ critères de recherche pertinents pour les utilisateurs
- L'association automatique entre les deux est à ce jour possible uniquement dans des domaines très restreints
- Bon moyen d'identifier la cible d'un utilisateur lors d'une session de recherche : **inclure l'utilisateur dans la boucle**
- Recherche **itérative** avec contrôle de pertinence (*relevance feedback*, voir aussi [CFB04])
- Peut être utilisé pour la recherche de textes [Sal68], musique, etc. où un fossé sémantique est également présent

21 janvier 2014

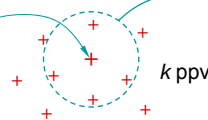
Bases de données multimédia

3

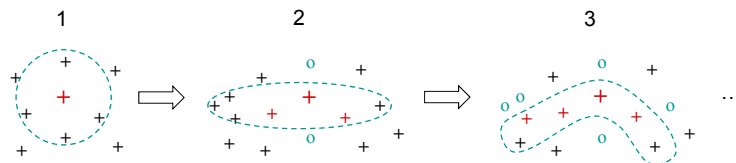
## Contrôle de pertinence : objectif

- Objectif : trouver des images sans employer de critère explicite dans les requêtes

### 1. Recherche par l'exemple



### 2. Recherche itérative avec contrôle de pertinence



21 janvier 2014

Bases de données multimédia

4

le cnam


## Contrôle de pertinence : exemple

Objectif : retrouver des portraits

Base de 7500 images, dont 110 portraits

Disponible : description globale (couleur, texture, forme)

Première page de résultats après 4 itérations →



21 janvier 2014 Bases de données multimédia [FCB04/1] 5

le cnam

## Composantes du mécanisme

1. **Learner** : à partir de l'information disponible (notamment des exemples positifs et/ou négatifs), estimer l'ensemble d'images visé
2. **Sélecteur** : à partir de l'estimation produite par le *learner*, choisir les images que l'utilisateur doit marquer lors de l'itération suivante
3. Utilisateur : fournir à chaque itération le retour pour les images choisies par le sélecteur
  - Les évaluations sont souvent faites à l'aide d'une vérité terrain, en **émulant** l'utilisateur

21 janvier 2014 Bases de données multimédia 6

## Difficultés pour l'apprentissage

- **Très peu d'exemples** étiquetés : leur nombre est souvent inférieur au nombre de dimensions de l'espace de description !
- **Déséquilibre** important entre le nombre d'exemples positifs et le nombre d'exemples négatifs
- **Forme** potentiellement **complexe** de l'ensemble d'images visé, qui peut même présenter **plusieurs modes distants** dans l'espace de description
- L'interactivité exige un **temps de réponse très court**, à la fois pour le *learner* et pour le sélecteur

## Learners employés

- Machines à vecteurs support (*support vector machines*, SVM, [SS02])
  - ◆ Avantages dans le contexte du retour de pertinence
    - La fonction de décision associée permet à la fois la définition d'une frontière et le classement des images
    - Avec un large choix des noyaux, les SVM permettent une grande liberté dans la forme des classes (avec un contrôle par la régularisation)
    - D'autres sources d'information (en dehors des exemples) permettent de définir des noyaux appropriés (*kernel engineering*)
    - Apprentissage très rapide avec le nombre relativement limité d'exemples fournis par le contrôle de pertinence
    - Moindre sensibilité au déséquilibre entre exemples positifs et négatifs
  - ◆ Inconvénients
    - Par rapport aux noyaux de Parzen, absence de caractère incrémental (dans la formulation de base) et donc étape de sélection plus coûteuse

le cnam

## SVM et discrimination d'images

21 janvier 2014      Bases de données multimédia      9

le cnam

## Sélecteur : objectifs et critères

- Objectifs
  1. Retourner un maximum d'images pertinentes à l'utilisateur
  2. Maximiser le transfert d'information **utilisateur** → **système**
 → Une même méthode ne pouvant pas répondre aux deux objectifs, on peut se servir de **2** sélecteurs (un par objectif)
- Critères de sélection
  - ◆ « **Most positive** » (MP) : retourner les images les plus pertinentes suivant l'estimation actuelle faite par le *learner* (se focalise sur l'objectif 1) – critère classique le plus utilisé
  - ◆ « **Most informative** » (MI) : retourner les images qui permettent à l'utilisateur de fournir un maximum d'information sur sa cible (se focalise sur l'objectif 2) → **minimiser le nombre d'exemples**
  - ◆ Des critères mixtes sont possibles

21 janvier 2014      Bases de données multimédia      10

## Contrôle de pertinence : exemple (2)

Objectif : retrouver des régions représentant des villages

Base avec 24000 régions, dont 87 dans la classe

Disponible : description des régions (couleur, texture, formes à l'intérieur)

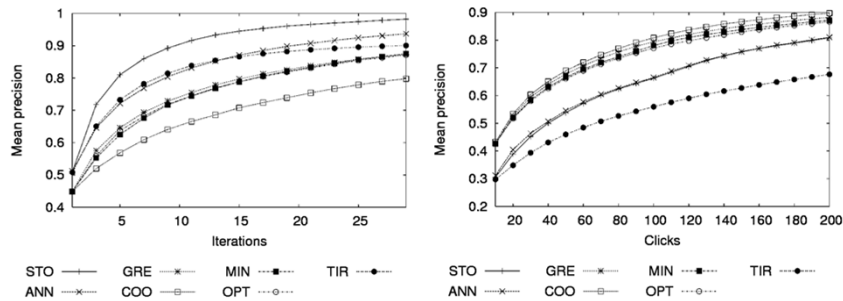
Première page de résultats avec 6 exemples positifs et 28 négatifs



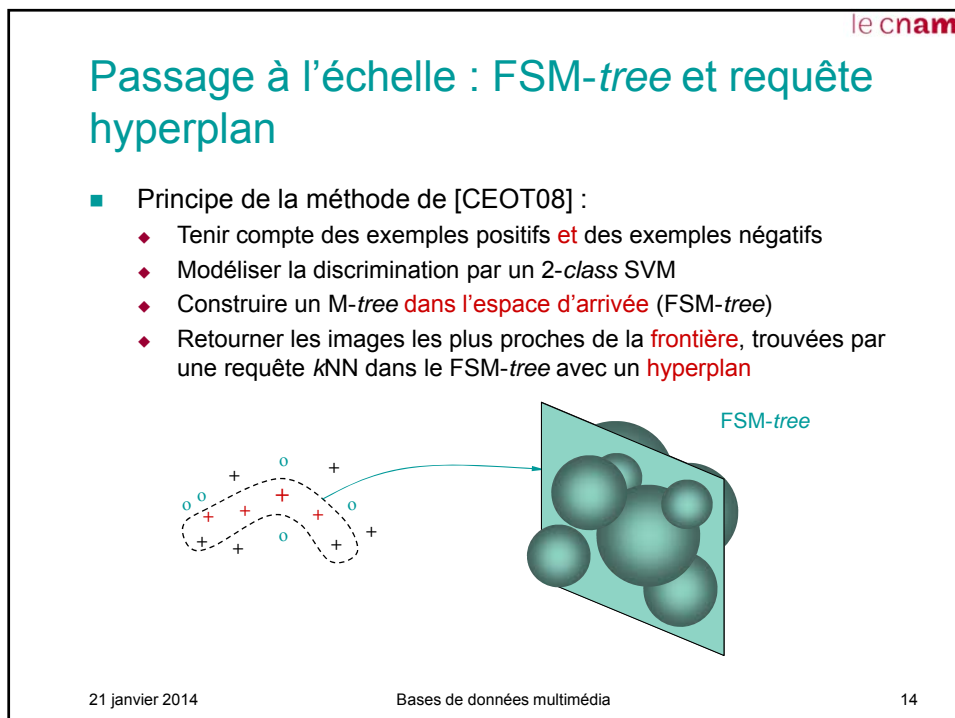
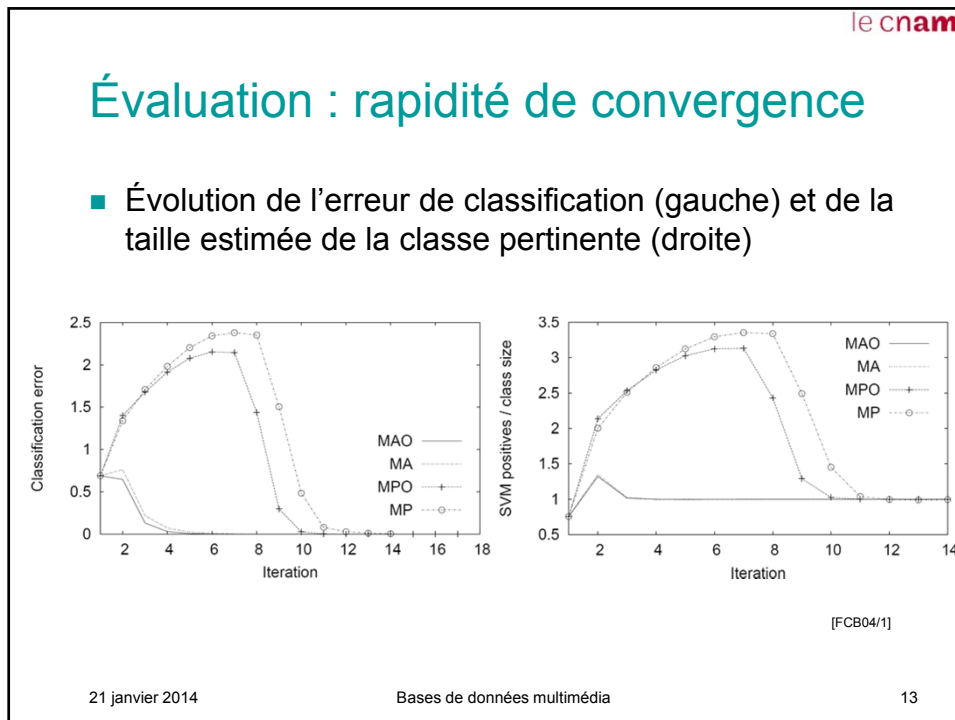
[FCB04/2]

## Évaluation : rapidité de convergence

- Évolution de la précision moyenne avec les itérations (gauche) ou les *clicks* (droite ; *click* = image marquée)



[CTF05]



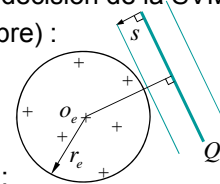
## FSM-tree et requête hyperplan (2)

■ Distance à l'hyperplan :  $d(Q, o_e) = \frac{|\sum_i \alpha_i y_i K(o_e, x_i) + b|}{\sqrt{\sum_{i,j} \alpha_i \alpha_j y_i y_j K(x_i, x_j)}}$

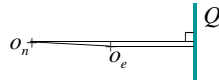
( $f(o_e) = \sum_i \alpha_i y_i K(o_e, x_i) + b$  étant la fonction de décision de la SVM)

- Principe d'élagage (test de rejet d'un sous-arbre) :

- ◆ Si  $d(Q, o_e) > r_e + s$  alors le nœud n'est pas conservé pour exploration ultérieure



- Comment éviter plus de calculs de distances :



$$d(Q, o_e) + d(o_e, o_n) \geq d(Q, o_n) \text{ mais } d(Q, o_e) + d(Q, o_n) \not\geq d(o_e, o_n)$$

$$\Rightarrow d(Q, o_e) \geq d(Q, o_n) - d(o_e, o_n) \quad (d(Q, o_e) \not\geq |d(Q, o_n) - d(o_e, o_n)|), \text{ donc}$$

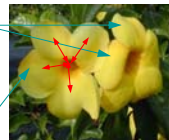
$$\text{si } d(Q, o_n) - d(o_e, o_n) > r_e + s \text{ alors } d(Q, o_e) > r_e + s$$

## Recherche pluri-modale : motivation

- Le contenu d'une base multimédia peut être caractérisé par des données de nature différente

- ◆ Descripteurs du contenu visuel
- ◆ Méta-données textuelles structurées (exemple : nom photographie, date, paramètres de la prise de vue)
- ◆ Méta-données textuelles non structurées ou semi-structurées : liste de mots-clés, textes descriptifs
- ◆ Graphes décrivant des relations spatiales entre composants

« Corolle glabre, jaune orange ... »



« Lobes obliquement ovales-orbiculaires »

## Types de mots-clés

1. Désignent des **objets identifiables** dans l'image ou des **caractéristiques visuelles** de ces objets
  - ◆ La détection et/ou la reconnaissance d'objets permet dans certains cas de les obtenir
2. Désignent des **caractéristiques visuelles de la scène**
  - ◆ La reconnaissance de scènes permet dans certains cas de les obtenir
3. Concernent l'**interprétation de la scène** (qu'un humain peut inférer)
  - ◆ Très difficiles à obtenir automatiquement à partir des images
  - ◆ Complètent le contenu visuel des images
4. Mettent l'image dans un **contexte** qui ne peut pas être inféré par un non spécialiste à partir du contenu visuel de l'image
  - ◆ Ne peuvent donc pas être obtenus automatiquement
  - ◆ Complètent le contenu visuel de l'image

## Image et mots-clés : exemple



© <http://la-revolution-des-oeillets.france.com/>

1. Char (/ tank), voiture, immeuble, personnes
2. Extérieur, jour, ville
3. Fraternalisation
4. Lisbonne, révolution des œillets, 25 avril 1974 , Caetano, Salazar

## Prise en compte des mots

- Objectif : mesurer la similarité entre les ensembles de mots associés à 2 images
- Sources
  1. Ensembles de mots-clés : mots pertinents mais peu nombreux
    - Exemple : « *sunset, seascape, sailboat* »
  2. Textes : nombreux mots, lesquels sont pertinents ?
    - Exemple : « Tôt le 25 avril 1974, au Portugal, des capitaines en rupture avec le système de Salazar se révoltent et prennent le pouvoir. La voix calme d'un mystérieux « Commandement du Mouvement des Forces armées » transmise par les radios de Lisbonne, Renascenta et Radio Clube donnant le signal de la révolte aux capitaines mutins, exhorte les gens à rester chez eux et à garder leur calme. »

## Prise en compte des mots (2)

- Sous quelle forme les prendre en compte ?
  1. Mots : forme directement disponible, mais relation complexe avec la signification (problèmes : homonymie, synonymie, etc.)
  2. Concepts : signification claire, mais le plus souvent forme non directement disponible (mot → concept : désambiguïsation...)
- Mesurer la similarité entre 2 ensembles de mots :
  1. Compter les **cooccurrences de mots** entre les 2 ensembles ; simple, mais traite les mots de façon égalitaire et ne tient pas compte d'éventuelles relations entre mots différents
  2. Relations issues d'**analyse statistique** : pondération différenciée des mots selon leur potentiel discriminatif, prise en compte des cooccurrences dans l'ensemble du corpus pour identifier une sémantique « latente » (⇒ désambiguïsation implicite)
  3. **Relations conceptuelles** : partie ou ensemble des relations issues d'une ontologie (*sort\_of, part\_of*, proximité conceptuelle, etc.)

le cnam

## Modèle vectoriel d'un texte

- Extraction des **termes** par l'élimination des mots peu spécifiques (conjonctions, prépositions, verbes auxiliaires, etc.) et lemmatisation  
« Tôt le 25 avril 1974, au Portugal, des capitaines en rupture avec le système de Salazar se révoltent et prennent le pouvoir. La voix calme d'un mystérieux « Commandement du Mouvement des Forces armées » transmise par les radios de Lisbonne, Rénascença et Radio Clube donnant le signal de la révolte aux capitaines mutins, exhorte les gens à rester chez eux et à garder leur calme. »
- Pour un ensemble donné de textes, association d'une dimension à chaque terme et représentation de chaque texte par le vecteur (de très grande dimension) correspondant aux termes qu'il contient [Sal68]
- Pondération des termes : nombreuses propositions, comme *tf\*idf* (term frequency \* inverse document frequency [SMG83]) :
 
$$\sum_k n_k \log \frac{|D|}{|\{d : t_i \in d\}|}$$

	armée	calme Clube	garder gens	pouvoir
1	0 0,5	0 0,5 0,4	0,3 0,2 0	0 0,3 0
10...4				

21 janvier 2014
Bases de données multimédia
21

le cnam

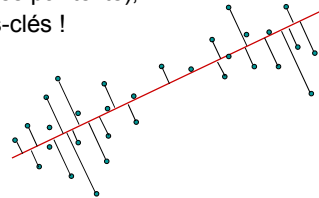
## Modèle vectoriel d'un texte (2)

- Mesurer la dissimilarité : cosinus de l'angle, distance L1 ou L2,...
- Quelques insuffisances du modèle vectoriel de base
  - ◆ Les relations entre mots (synonymie, antonymie, hyperonymie) ne sont pas représentées
    - Par exemple, si les termes « voiture », « automobile » et « véhicule » sont rencontrés dans les textes, 1 dimension différente est affectée à chacun
  - ◆ L'ambiguïté (pas uniquement due à l'homonymie) n'est pas traitée
  - ◆ Grand nombre de mots peu fréquents, qui ne sont pas pertinents pour autant → présence de « bruit »
- L'impact de ces insuffisances augmente quand le nombre de termes par texte diminue, or les ensembles de mots clés associés aux images sont en général de faible cardinalité !

21 janvier 2014
Bases de données multimédia
22

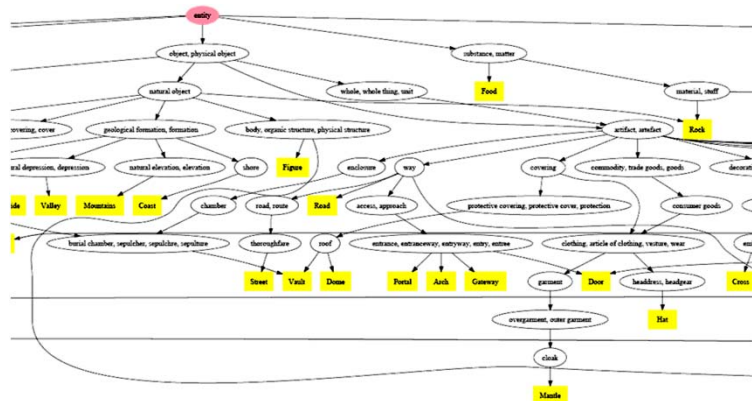
## Relations statistiques

- (analyse sémantique latente, *latent semantic analysis*, LSA, [DDF90])
- Analyse en composantes principales (ACP) sur l'ensemble des vecteurs qui représentent les textes d'un corpus (modèle vectoriel), projection sur les composantes de **plus forte variance**
- Éliminer le « bruit » qui domine les composantes de faible variance
- Mettre en évidence des « composantes sémantiques cachées » (résultant de cooccurrences) qui représentent des relations **statistiquement déterminées** entre mots et à traiter partiellement l'ambiguïté
- Condition : les caractéristiques pertinentes doivent être bien représentées (→ beaucoup de textes, beaucoup de termes par texte), peu probable avec des ensembles de mots-clés !



## Relations conceptuelles

- Relations issues d'une **ontologie** (*a priori*, non liées à un corpus)
- Fragment de WordNet



[Fer04], <http://www-rocq.inria.fr/media>

le cnam

## Relations conceptuelles (2)

- Avantages
  - ◆ Fiables quel que soit le volume du corpus
  - ◆ Relative complétude par rapport aux relations possibles
- Désavantages
  - ◆ Étape de désambiguïsation préalable
  - ◆ Non prise en compte des spécificités du corpus
  - ◆ Insuffisances de l'ontologie

```

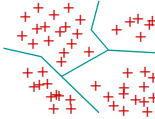

graph TD
    entity((entity)) --> object([object, physical object])
    object --> whole([whole, whole thing, unit])
    whole --> artifact([artifact, artefact])
    artifact --> creation([creation])
    creation --> art([art, fine art])
    art --> plastic([plastic art])
    plastic --> sculpture((sculpture))
    
    abstraction((abstraction)) --> attribute([attribute])
    attribute --> shape([shape, form])
    shape --> figure([figure])
    figure --> solid([solid figure, three-dimensional figure])
    solid --> sculpture
    
    sculpture --> statue[statue]
    
    artifact --> sculpture
    creation --> sculpture
    art --> sculpture
    plastic --> sculpture
    
    sculpture --- relations["(relations sort_of uniquement)"]
  
```

21 janvier 2014 Bases de données multimédia [Fer05] 25

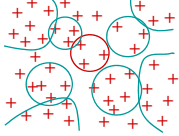
le cnam

## Catégorisation du contenu des bases

- Rendre explicite une structure qui peut être mise en rapport avec des critères de recherche des utilisateurs
- Partitionnement de la collection
  - ◆ Résumés globaux, *a priori*
    - Quelle granularité ?
    - Quelle pertinence ?
    - Niveau sémantique ?
  - ◆ Vues locales, contextuelles (problème peu abordé)
    - Structurer les alternatives à chaque étape de la recherche
    - Niveau sémantique ?

© <http://arabidopsis.info>



21 janvier 2014 Bases de données multimédia 26

## Classification automatique

(*cluster analysis, clustering*)

- Objectif général : répartir un ensemble donné d'individus en **groupes** (catégories, classes, taxons, *clusters*) de façon à regrouper les individus similaires et à séparer les individus dissimilaires
- Aucune information n'est disponible concernant l'appartenance de certaines données à certaines classes ; le nombre de groupes recherchés peut être connu *a priori* ou non
- Utilisations
  - ◆ Mettre en évidence une **structure** (simple) dans les données
  - ◆ Résumer les données par les **représentants** des groupes

## Distinctions entre les méthodes

- Par l'objectif
  - ◆ Partitionnement
  - ◆ Hiérarchie de groupes (→ plusieurs partitionnements disponibles)
- Par la nature des données : numériques, catégorielles, mixtes
- Par la représentation des données
  - ◆ Riche : représentation vectorielle permettant de définir centre de gravité, densité, intervalles, différentes métriques
  - ◆ Simple : l'unique structure disponible est une métrique
- Par la nature des groupes recherchés : mutuellement exclusifs ou non, nets ou flous
- Par la définition des groupes (critère de regroupement)
  - ◆ Ensembles **compacts** éloignés entre eux
  - ◆ Ensembles **denses** séparés par des régions moins denses

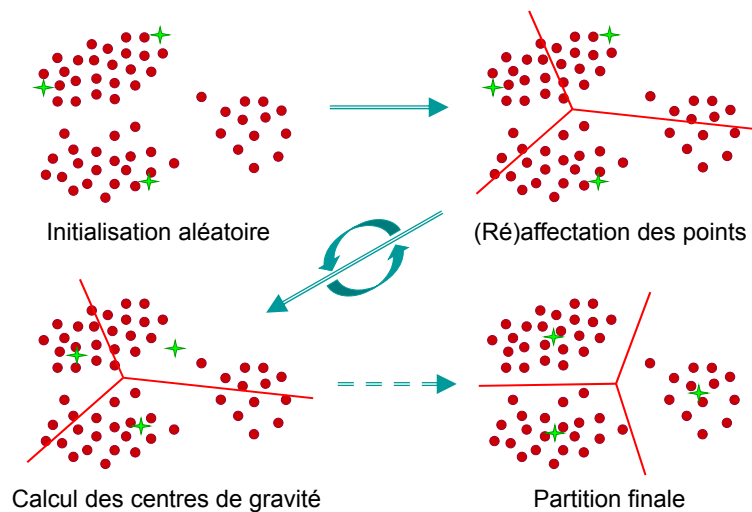
## Méthode des centres mobiles

- Objectif : répartir les  $n$  individus en  $k$  groupes disjoints  $\mathcal{E}_1, \mathcal{E}_2, \dots, \mathcal{E}_k$  (inconnus *a priori*) en optimisant une fonctionnelle de coût
- Données : ensemble  $\mathcal{E}$  de  $n$  individus décrits par  $p$  variables à valeurs dans  $\mathbb{R}$
- Distance dans  $\mathbb{R}^p$  (euclidienne usuelle, Mahalanobis, Minkowsky, etc.) qui définit la nature des (dis)similarités
- Fonctionnelle à optimiser ( $\mathbf{m}_j$  : centre de gravité du groupe  $\mathcal{E}_j$ ) :

$$J(\mathbf{M}) = \sum_{j=1}^k \sum_{\mathbf{x}_i \in \mathcal{E}_j} d^2(\mathbf{x}_i, \mathbf{m}_j) \quad (= \text{somme inerties intra-classe})$$

- Principe de la méthode : à partir d'une initialisation aléatoire de  $k$  vecteurs  $\mathbf{m}_j$ , alterner entre une étape de (ré)affectation des individus aux groupes (à partir de leur distances aux différents  $\mathbf{m}_j$ ) et une étape de (re)calcul des centres de gravité

## Centres mobiles : algorithme



## Centres mobiles : algorithme (2)

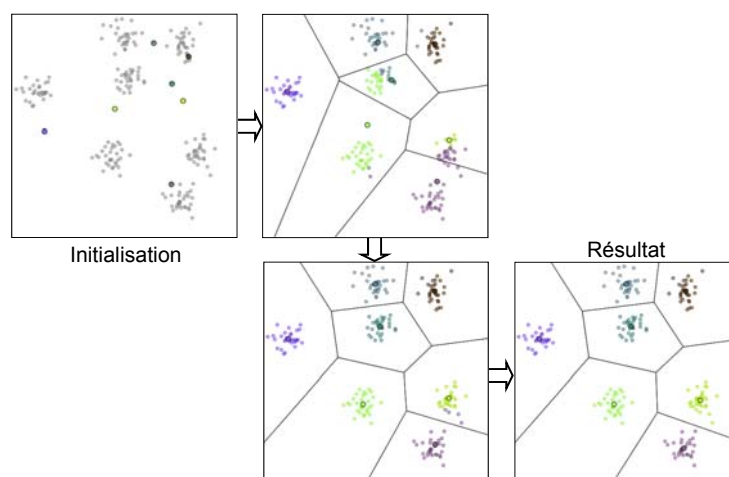
1. Initialisation aléatoire des centres (dans le domaine couvert par les  $n$  individus)
2. Tant que les centres ne sont pas stabilisés, répéter
  - ◆ Association de chaque individu au centre dont il est le plus proche (suivant la distance  $d$ )
  - ◆ Calcul du centre de gravité de l'ensemble d'individus associés au même centre et remplacement des anciens centres par les nouveaux
3. Chaque centre définit un groupe : l'ensemble des individus les plus proches de ce centre
  - On peut montrer que l'algorithme converge vers un minimum de la fonctionnelle, mais la fonctionnelle possède en général de nombreux minima locaux

21 janvier 2014

Bases de données multimédia

31

## Centres mobiles : illustration

Démonstration : [www.know-center.tugraz.at/download\\_extern/WE/clustering/kmeansdemo.jnlp](http://www.know-center.tugraz.at/download_extern/WE/clustering/kmeansdemo.jnlp)

21 janvier 2014

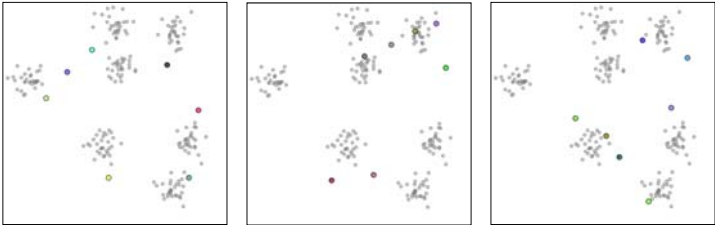
Bases de données multimédia

32

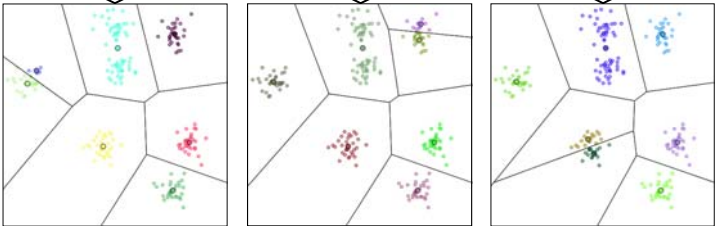
le cnam

## Centres mobiles : illustration (2)

Initialisation



Résultat



Démo : [www.know-center.tugraz.at/download\\_external/WE/clustering/kmeansdemo.inlp](http://www.know-center.tugraz.at/download_external/WE/clustering/kmeansdemo.inlp)

21 janvier 2014      Bases de données multimédia      33

le cnam

## Centres mobiles : intérêt et limitations

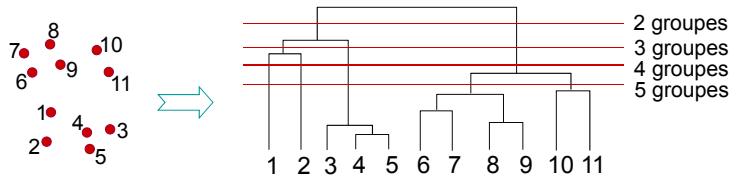
- Intérêt :
  - ◆ Simplicité...
  - ◆ Paramètre unique : valeur souhaitée pour le nombre de classes
  - ◆ Faible complexité :  $O(t \cdot k \cdot n)$  ( $t$  étant le nombre d'itérations)
- Limitations et solutions :
  - ◆ Données vectorielles uniquement (pour calculer les moyennes) : limitation levée dans des méthodes dérivées (ex. *k-medoids*)
  - ◆ Classes de forme sphérique (si la distance euclidienne usuelle est employée) : on peut se servir de la distance de *Mahalanobis* (calculée par classe)
  - ◆ Dépendance des conditions initiales (car convergence vers minimum local) : diverses solutions (notamment stochastiques)
  - ◆ Sensibilité au bruit : utilisation d'une fonctionnelle de coût **robuste**, estimation **robuste** des moyennes
  - ◆ Choix *a priori* difficile du nombre de classes : introduction d'un terme de **régularisation** dans la fonctionnelle de coût

21 janvier 2014      Bases de données multimédia      34

le cnam

## Classification hiérarchique

- Objectif : obtenir une **hiérarchie de groupes**, qui offre une structure plus riche qu'un simple partitionnement (et permet d'obtenir plusieurs partitionnements à différents niveaux de granularité)



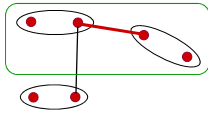
- Méthodes :
  - ◆ Ascendantes : procèdent par agglomération (agrégation) d'individus et de groupes ;
  - ◆ Descendantes : procèdent par partitionnement de l'ensemble initial, puis des groupes

21 janvier 2014
Bases de données multimédia
35

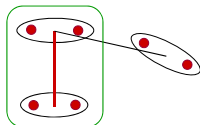
le cnam

## Classification ascendante hiérarchique

- Procède par agrégation, à chaque étape, des 2 groupes (chaque individu étant également considéré comme un groupe) qui sont les plus proches (les moins dissimilaires)
- Mesures de dissimilarité entre 2 groupes (indices d'agrégation) :
  - ◆ Indice du lien minimum (*single linkage*) :
 



$$\delta(h_p, h_q) = \min_{\substack{\mathbf{x}_i \in h_p \\ \mathbf{x}_j \in h_q}} d(\mathbf{x}_i, \mathbf{x}_j)$$
  - ◆ Indice de Ward (augmentation de l'inertie de la réunion des groupes) :
 



$$\delta(h_p, h_q) = \frac{|h_p| |h_q|}{|h_p| + |h_q|} d^2(\mathbf{m}_p, \mathbf{m}_q)$$

$\mathbf{m}_p, \mathbf{m}_q$  : centres de gravité des groupes

21 janvier 2014
Bases de données multimédia
36

## CAH : algorithme

- Remarque : certains indices d'agrégation exigent des données vectorielles, pour d'autres un espace métrique suffit
  - Choix nécessaires : mesure de distance entre individus, mesure de dissimilarité entre les groupes (indice d'agrégation)
  - Algorithme (complexité  $O(n^2 \log n)$ ) :
    1. Initialisation : nombre de groupes au départ ( $N_g$ ) = nombre d'individus ( $n$ )
    2. Répéter
      - ◆ Calcul de l'indice d'agrégation entre toutes les paires de  $N_g$  groupes issues de l'itération précédente
      - ◆ Regroupement des 2 groupes les plus proches (valeur la plus petite de l'indice d'agrégation); les autres groupes restent inchangés ;  
 $N_g := N_g - 1$
- tant que  $N_g > 1$

## Validité de la classification

- **Convergence** vers un résultat **quelles que soient les données**, mais **quelle est la validité de ce résultat** ?
- Principales questions
  1. Y a-t-il réellement des regroupements « naturels » dans les données ?
  2. Les groupes identifiés sont-ils en accord avec nos (éventuelles) connaissances *a priori* du problème (validation **externe**) ?
  3. Les groupes identifiés sont-ils bien « ajustés » aux données (validation **interne**) ? Indices proposés : statistique modifiée de Hubert (alignement distance  $\leftrightarrow$  partition), indice Davies-Bouldin (rapport des inerties), indice de Dunn (rapport diamètre/distance)...
  4. Les résultats de la méthode A sont-ils meilleurs que les résultats de la méthode B (validation **relative**) ? Application des indices définis pour la validation interne, possibilités de sélection de modèle.

## Classification semi-supervisée

1. Similarités issues des données décrivant les images (descripteurs visuels, mots-clés, relations spatiales...)
    - ◆ Directement calculables là où les données correspondantes sont disponibles
    - ◆ Nombreuses et bon marché
  2. Similarités fournies par les utilisateurs
    - ◆ Explicitement ou à travers des corrélations de recherche, ...
    - ◆ Étiquettes de classe, contraintes, ...
    - ◆ Caractère approximatif, domaine partiel de définition
    - ◆ Plus rares et coûteuses
- **Combiner ces sources** d'information

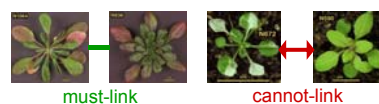
21 janvier 2014

Bases de données multimédia

39

## Exemple : similarités & contraintes [GCB08]

- Situation visée (catégorisation de bases d'images) :
  - ◆ Grande base inconnue (ou peu connue)
  - ◆ Classification simple : résultats médiocres
    - Supervision nécessaire !
  - ◆ Classes d'images inconnues *a priori*
    - L'utilisateur ne peut pas donner d'étiquettes mais est capable de dire si 2 images devraient être dans une même classe (contrainte *must-link*) ou dans des classes différentes (contrainte *cannot-link*)
    - Étant donnée la taille de la base, la quantité de supervision (contraintes) doit être minimale



21 janvier 2014

Bases de données multimédia

40

## Références

- [BH01] A. Budanitsky, G. Hirst. Semantic distances in WordNet: an experimental, application-oriented evaluation of five measures. In *Proceedings of NAACL 2001*.
- [CEOT08] Crucianu, M., Estevez, D., Oria, V., Tarel, J.-Ph. Speeding Up Active Relevance Feedback with Approximate kNN Retrieval for Hyperplane Queries, *International Journal of Imaging Systems and Technology*, Vol. 18 (2-3): 150-159, 2008.
- [CFB04] Crucianu, M., Ferecatu, M., Boujemaa, N. Relevance feedback for image retrieval: a short survey, juin 2004, 20 p., dans *State of the Art in Audiovisual Content-Based Retrieval, Information Universal Access and Interaction, Including Datamodels and Languages*, rapport du Réseau d'Excellence DELOS2 (6ePCRD).
- [CTF05] Crucianu, M., Tarel, J.-Ph., Ferecatu, M. A Comparison of User Strategies in Image Retrieval with Relevance Feedback, *7th International Workshop on Audio-Visual Content and Information Visualization in Digital Libraries (AVIDiLib'05)*, pp. 121-130, Cortona, Italie, 2005.
- [FBC05] M. Ferecatu, N. Boujemaa, M. Crucianu. Hybrid Visual and Conceptual Image Representation in an Active Relevance Feedback Context. *Proceedings of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval*, Singapour, 2005.
- [FCB04/1] Ferecatu, M., Crucianu, M., Boujemaa, N. Retrieval of Difficult Image Classes Using SVM-Based Relevance Feedback, *Proceedings of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval*, 2004, New York, pp. 23-30.
- [Fer05] Ferecatu, M. Image retrieval with active relevance feedback using both visual and keyword-based descriptors, *Thèse de doctorat*, Université de Versailles Saint-Quentin-en-Yvelines, 2005.

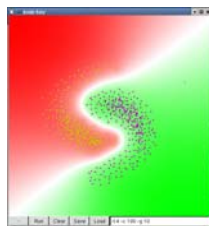
## Références

- [FGB05] Y. Fang, D. Geman and N. Boujemaa. An Interactive System for Mental Face Retrieval, *7th ACM SIGMM International Workshop on Multimedia Information Retrieval (MIR'05)*, November 10-11, 2005, Singapore.
- [FK97] H. Frigui, R. Krishnapuram. Clustering by competitive agglomeration. *Pattern Recognition* 30 (7): 1109-1119, 1997.
- [FS03] F. Fleuret, H. Sahbi. Scale-invariance of support vector machines based on the triangular kernel. In *3rd Intl. Workshop on Statistical and Computational Theories of Vision*, October 2003.
- [GCB08] N. Grira, M. Crucianu, N. Boujemaa. Active Semi-Supervised Fuzzy Clustering, *Pattern Recognition*, Vol. 41, No 5, pp. 1851-1861.
- [JMF99] A. Jain, M. Murty, P. Flynn. Data clustering: a review. *ACM Computing Surveys*, 31(3):264-323, 1999.
- [Sal68] G. Salton. *Automatic Information Organization and Retrieval*. McGraw-Hill, 1968.
- [Sch00] B. Schölkopf. The kernel trick for distances. In *Advances in Neural Information Processing Systems*, volume 12, pages 301-307. MIT Press, 2000.
- [SMG83] G. Salton and M. J. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, 1983.
- [SS03] B. Schölkopf, A. Smola. *Learning with Kernels*. MIT Press, 2002.
- [TK00] S. Tong, D. Koller. Support vector machine active learning with applications to text classification. In *Proceedings of ICML-00, 17th International Conference on Machine Learning*, pp. 999-1006. Morgan Kaufmann, 2000.

## Annexe 1 : machines à vecteurs support

### ■ Support Vector Machines

1. Séparateurs à Vastes Marges
  - ◆ SVM ne servent pas seulement à séparer (aussi à la régression, ...)
  - ◆ SVM pas les seuls séparateurs à vastes marges (autre ex. : *boosting*)
2. Machines à vecteurs support



Exemple simple avec noyau angulaire  
 Intensité de la couleur proportionnelle à l'éloignement de la frontière  
 (outil employé : version maison de svm-toy)

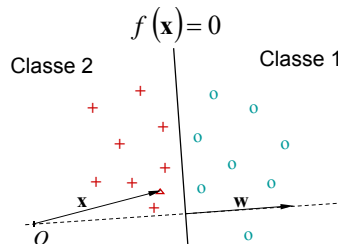
21 janvier 2014

Bases de données multimédia

43

## SVM et séparation linéaire

- Données d'apprentissage  $D_n = \{(\mathbf{x}_i, y_i)\}_{1 \leq i \leq n}$ ,  $y_i \in \{-1, +1\}$
- On cherche une fonction de décision  $f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b$ , basée sur l'hyperplan  $\mathbf{w}^T \mathbf{x} + b = 0$ ,  $\mathbf{w}$  étant un vecteur normal à l'hyperplan
  - ◆ Affectation à la classe 1 si  $f(\mathbf{x}) > 0$
  - ◆ Affectation à la classe 2 si  $f(\mathbf{x}) < 0$
- Séparabilité linéaire :  $\exists \mathbf{w}, b$  tels que  $y_i f(\mathbf{x}_i) > 0$  pour  $1 \leq i \leq n$
- Remarque : si la condition est valable pour  $\mathbf{w}, b$ , alors elle est valable pour  $k\mathbf{w}, kb, \forall k > 0$



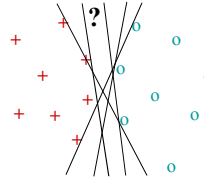
21 janvier 2014

Bases de données multimédia

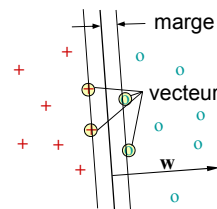
44

## Discrimination linéaire et marge

- Quelle séparation choisir quand plusieurs sont possibles ?



- Une possibilité : choisir la séparation qui maximise la **marge** (distance minimale entre un exemple et la surface de séparation)



en fonction de  $\mathbf{w}$  :  $\text{marge} = \frac{|\mathbf{x}_s^T \mathbf{w} + b|}{\|\mathbf{w}\|}$

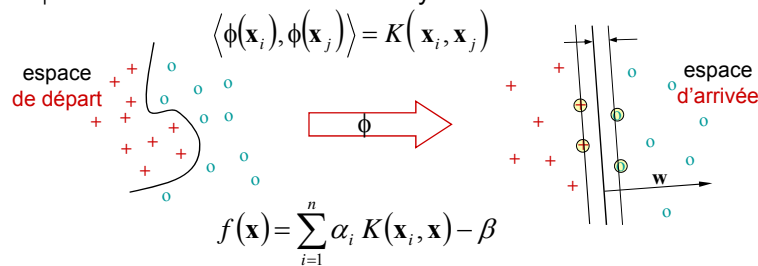
normalisation :  $|\mathbf{x}_s^T \mathbf{w} + b| = 1$  pour tout

**vecteur support**  $\mathbf{x}_s$

donc  $\text{marge} = 1/\|\mathbf{w}\|$

## SVM : astuce des noyaux

- Comment étendre ces résultats à des séparateurs non linéaires ?
- Principe : transposer les données dans un autre espace (en général de plus grande dimension) dans lequel elles sont linéairement séparables (ou presque) ← *kernel trick*
- Transformation  $\phi: \mathbb{R}^d \rightarrow \mathcal{H}$ ,  $\mathbf{x} \rightarrow \phi(\mathbf{x})$  ( $\mathcal{H}$  espace de Hilbert)
- Sous certaines conditions, l'existence de  $\phi$  et de  $\mathcal{H}$  est garantie et  $\phi$  est associée à une fonction noyau  $K: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$  :

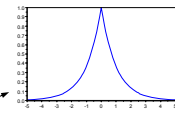
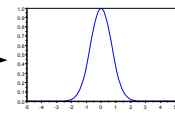
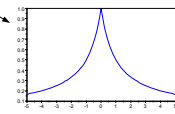
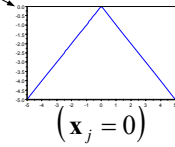


$$\langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle = K(\mathbf{x}_i, \mathbf{x}_j)$$

$$f(\mathbf{x}) = \sum_{i=1}^n \alpha_i K(\mathbf{x}_i, \mathbf{x}) - \beta$$

le cnam

## SVM : exemples de noyaux

- Linéaire :  $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j = \langle \mathbf{x}_i, \mathbf{x}_j \rangle$
- Exponentiel :  $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|)$  
- Gaussien (RBF) :  $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2)$  
- Hyperbolique :  $K(\mathbf{x}_i, \mathbf{x}_j) = 1 / (\varepsilon + \gamma \|\mathbf{x}_i - \mathbf{x}_j\|)$  
- Angulaire :  $K(\mathbf{x}_i, \mathbf{x}_j) = -\|\mathbf{x}_i - \mathbf{x}_j\|$ 
  - ◆ Noyau **conditionnellement** positif défini [Sch00]
  - ◆ Engendre une **invariance à l'échelle** [FS03]
- Puissance :  $K(\mathbf{x}_i, \mathbf{x}_j) = -\|\mathbf{x}_i - \mathbf{x}_j\|^\beta$   
( $0 < \beta \leq 2$ ) 

(x<sub>j</sub> = 0)

21 janvier 2014 Bases de données multimédia 47

le cnam

## Annexe 2 : classification automatique floue

- Objectif : répartir les  $n$  individus en  $k$  groupes flous  $\mathcal{E}_1, \mathcal{E}_2, \dots, \mathcal{E}_k$  (inconnus *a priori*) en optimisant une fonctionnelle de coût
- Groupes flous (*fuzzy*) : frontières ne sont pas nettes, un même élément appartiendra à différents **degrés** à plusieurs groupes
- Données : ensemble  $\mathcal{E}$  de  $n$  individus décrits par  $p$  variables à valeurs dans  $\mathbb{R}$  (des extensions à l'utilisation de *medoids* existent)
- Fonctionnelle à optimiser :
 
$$J(\mathbf{M}) = \sum_{j=1}^k \sum_{i=1}^n u_{ij}^2 d^2(\mathbf{x}_i, \mathbf{m}_j)$$
 où le centre de gravité du groupe  $\mathcal{E}_j$  est  $\mathbf{m}_j = \frac{\sum_{i=1}^n u_{ij}^2 \mathbf{x}_i}{\sum_{i=1}^n u_{ij}^2}$  et
 
$$u_{ij} = \frac{1/d^2(\mathbf{x}_i, \mathbf{m}_j)}{\sum_{q=1}^k 1/d^2(\mathbf{x}_i, \mathbf{m}_q)}$$
 est le **degré d'appartenance** de  $\mathbf{x}_i$  au groupe  $\mathcal{E}_j$ , avec la contrainte de normalisation  $\sum_{j=1}^k u_{ij} = 1$

21 janvier 2014 Bases de données multimédia 48

## C-moyennes floues : algorithme

- Algorithme de classification automatique floue d'individus situés dans  $\mathbb{R}^p$ ;  $\mathbf{m}_j$  sont les centres de gravité des groupes
- Algorithme (obtenu par minimisation sous contraintes du coût) :
  1. Initialisation aléatoire des centres (dans le domaine couvert par les  $n$  individus)
  2. Répéter
    - ◆ Calcul du degré d'appartenance de chaque individu à chaque groupe
    - ◆ Calcul du centre de gravité de l'ensemble d'individus du même groupe et remplacement des anciens centres par les nouveaux
 tant que les centres ne sont pas stabilisés
  3. Chaque groupe contient soit tous les individus mais avec les degrés d'appartenance obtenus (groupes flous), soit l'ensemble des individus les plus proches d'un centre (groupes nets, *crisp*)

## Agglomération compétitive

(*competitive agglomeration*, CA [FK97])

- Objectif : introduire une composante de **régularisation** dans la fonction de coût, permettant à l'utilisateur de **ne pas fixer a priori le nombre de groupes**
- Fonctionnelle à optimiser :
 
$$J(\mathbf{M}) = \sum_{j=1}^k \sum_{i=1}^n u_{ij}^2 d^2(\mathbf{x}_i, \mathbf{m}_j) - \beta \sum_{j=1}^k \left( \sum_{i=1}^n u_{ij} \right)^2$$

$\beta$  étant un paramètre qui décroît de façon exponentielle au fil des itérations (engendre une « **compétition** » entre les groupes)
- Les degrés d'appartenance sont actualisés suivant

$$u_{ij} = \underbrace{\frac{1/d^2(\mathbf{x}_i, \mathbf{m}_j)}{\sum_{q=1}^k 1/d^2(\mathbf{x}_i, \mathbf{m}_q)}}_{u_{ij}^{\text{FCM}}} + \underbrace{\frac{\beta}{d^2(\mathbf{x}_i, \mathbf{m}_j)}}_{u_{ij}^{\text{Biais}}} \left[ N_j - \frac{\sum_{q=1}^k N_q / d^2(\mathbf{x}_i, \mathbf{m}_q)}{\sum_{q=1}^k 1/d^2(\mathbf{x}_i, \mathbf{m}_q)} \right]$$

## Agglomération compétitive (2)

- Les **effectifs flous** des groupes,  $N_j$ , sont donnés par  $N_j = \sum_{i=1}^n u_{ij}$  et l'évolution de  $\beta$  au fil des itérations  $t$  est

$$\beta(t) = \eta_0 e^{-t/\tau} \cdot \frac{\sum_{j=1}^k \sum_{i=1}^n u_{ij}^2 d^2(\mathbf{x}_i, \mathbf{m}_j)}{\sum_{j=1}^k \left( \sum_{i=1}^n u_{ij} \right)^2}$$

- Suite à la compétition, les groupes dont l'effectif flou devient inférieur à un seuil (en général choisi très bas) sont éliminés, ainsi que leurs centres
- Le paramètre « **nombre de groupes** ( $k$ ) », difficile à choisir, est donc remplacé par la **valeur initiale** et la **vitesse de décroissance** de  $\beta$ , auxquelles l'algorithme est moins sensible (les mêmes résultats sont obtenus pour une plage assez large de valeurs), plus le **seuil d'élimination**, défini à partir de la résolution souhaitée

## CA : algorithmes

- Algorithme (obtenu par minimisation sous contraintes du coût) :
  1. Initialisation aléatoire des centres (dans le domaine couvert par les  $n$  individus)
  2. Répéter
    - ◆ Calcul de  $\beta$  pour l'itération courante
    - ◆ Calcul du degré d'appartenance de chaque individu à chaque groupe
    - ◆ Calcul du centre de gravité de l'ensemble d'individus du même groupe et remplacement des anciens centres par les nouveaux
    - ◆ Calcul de l'effectif flou de chaque groupe et élimination des groupes dont l'effectif est inférieur à un seuil (et des centres correspondants)
 tant que les centres ne sont pas stabilisés
  3. Chaque groupe contient soit tous les individus mais avec les degrés d'appartenance obtenus (groupes flous), soit l'ensemble des individus les plus proches d'un centre (groupes nets, *crisp*)

## Intérêt et limitations

- C-moyennes floues :
  - ◆ Avantages : conserve la complexité des centres mobiles, tout en étant moins sensible au bruit grâce à la pondération par les degrés d'appartenance
  - ◆ Limitations : mêmes que pour centres mobiles (hors sensibilité au bruit)
- Agglomération compétitive :
  - ◆ Avantages :
    - Moins sensible à l'initialisation : l'utilisation de nombreux centres assure leur distribution plus homogène dans les données
    - Ne demande pas à l'utilisateur de fixer *a priori* le « bon » nombre de groupes (dans la pratique, l'algorithme fonctionne bien quand le nombre de groupes est initialisé à une valeur élevée)
  - ◆ Limitations : l'automatisation de la classification reste incomplète...
- Ces méthodes peuvent être étendues à l'utilisation de *medoids*