

OKM : Une librairie R pour la classification recouvrante

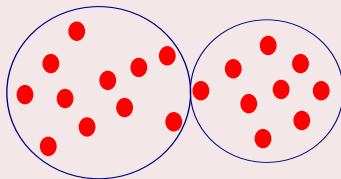
Guillaume Cleuziou, Léo Rousseau

Université d'Orléans - LIFO

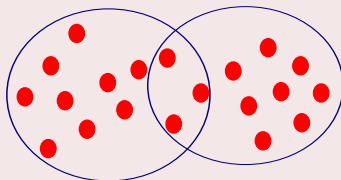
2 Juillet 2012

Problématique

- Classification non supervisée (clustering) stricte



- Classification recouvrante



Motivation de la classification recouvrante

- Recherche d'informations (textes, images, vidéo)
- Réseaux sociaux
- Bio-informatique

Algorithmes

k-means [McQueen 1967]

Critère objectif :

$$\mathcal{J}(\{C_k\}_{k=1}^K) = \sum_{k=1}^K \sum_{x_i \in C_k} d^2(x_i, C_k)$$

Deux étapes

- Affectation
- Mise à jour des représentants

Algorithmes

k-means [McQueen 1967]

Critère objectif :

$$\mathcal{J}(\{C_k\}_{k=1}^K) = \sum_{k=1}^K \sum_{x_i \in C_k} d^2(x_i, C_k)$$

Deux étapes

- Affectation
- Mise à jour des représentants

okm [Cleuziou 2007]

Critère objectif :

$$\mathcal{J}(\{C_k\}_{k=1}^K) = \sum_{x_i \in X} d^2(x_i, \phi(x_i))$$

$\phi(x_i) \rightarrow$ Combinaison linéaire $\{c_k\}$

Algorithmes

k-means [McQueen 1967]

Critère objectif :

$$\mathcal{J}(\{C_k\}_{k=1}^K) = \sum_{k=1}^K \sum_{x_i \in C_k} d^2(x_i, C_k)$$

Deux étapes

- Affectation
- Mise à jour des représentants

okm [Cleuziou 2007]

Critère objectif :

$$\mathcal{J}(\{C_k\}_{k=1}^K) = \sum_{x_i \in X} d^2(x_i, \phi(x_i))$$

$\phi(x_i) \rightarrow$ Combinaison linéaire $\{c_k\}$

Deux étapes

- Affectation (Heuristique)
- Mise à jour des représentants

Algorithme WOKM

weighted k-means [Chan et al 2004]

Critère objectif :

$$\mathcal{J}(\{C_k\}_{k=1}^K) = \sum_{k=1}^K \sum_{x_i \in C_k} \sum_{v=1}^p \lambda_{k,v}^\beta |x_{i,v} - c_{k,v}|^2$$

Trois étapes

- Affectation
- Mise à jour des représentants
- Mise à jour des poids

wokm [Cleuziou 2009]

Critère objectif :

$$\mathcal{J}(\{C_k\}_{k=1}^K) = \sum_{x_i \in X} \sum_{v=1}^p \gamma_{i,v}^\beta |x_{i,v} - \phi_v(x_i)|^2$$

Trois étapes

- Affectation (Heuristique)
- Mise à jour des représentants
- Mise à jour des poids (Heuristique)

Librairie R : OKM

Fonctions disponibles

- `kmeans(X, centers, iter.max, nstart, ...)`
- `okm(X, centers, iter.max, nstart, visu, method)`
- `wokm(X, centers, iter.max, nstart, B, visu)`

Librairie R : OKM

Fonctions disponibles

- `kmeans(X, centers, iter.max, nstart, ...)`
- `okm(X, centers, iter.max, nstart, visu, method)`
- `wokm(X, centers, iter.max, nstart, B, visu)`

La classe okm

- Clusters
- Representatives
- Withiness
- Overlaps

Ex : Clusters

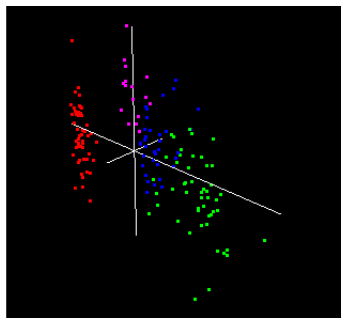
	1	2	3
1	0	1	0
2	0	1	1
3	0	1	1
4	0	0	1
5	1	0	0

Visualisations 3D

Objectif : *Appréhender le processus de construction des classes recouvrantes*

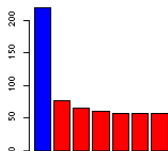
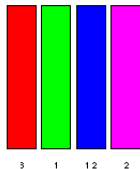
- Visualiser les individus
- Visualiser la formation des clusters
- Observer la décroissance du critère objectif

Outils de Visualisations 3D : Explorer la Classification

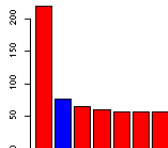
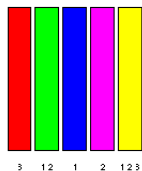
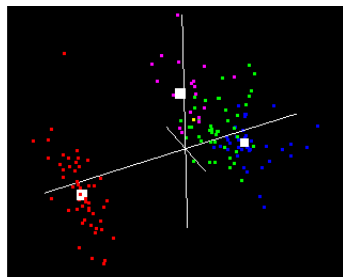


Commandes saisies

```
okm3d(iris[,-5], 3, 8)
```



Outils de Visualisations 3D : Explorer la Classification



Commandes saisies

- `okm3d(iris[-5], 3, 8)`
- `next.step()`
- `show.representatives()`
- `previous.step()`
- `hide.representatives()`
- ...

Evolutions de la Librairie

- Soumission du package au CRAN
- Extension à la norme L1
- Ajout de nouveaux algorithmes de classification recouvrante :
 - Model-based Overlapping Clustering [Banerjee 2005]
 - Pyramides [Diday 1984]
 - ...