

Classification de Données Fonctionnelles selon un Modèle Bayésien en Dimension Finie

Benjamin Auder

CEA - UPMC

28 mai 2009

Thèse CEA - UPMC

Directeur de thèse : Gérard Biau (UPMC)

Encadrant CEA : Bertrand Iooss (CEA)

Cadre d'étude

Code de calcul à *entrées vectorielles* et *sorties fonctionnelles*, **coûteux**.

$$\begin{pmatrix} x_{11} & \dots & x_{1p} \\ \vdots & \vdots & \vdots \\ x_{n1} & \dots & x_{np} \end{pmatrix} \longrightarrow \begin{pmatrix} y_1(t) \\ \vdots \\ y_n(t) \end{pmatrix} = \begin{pmatrix} y_1(t_1) & \dots & y_1(t_d) \\ \vdots & \vdots & \vdots \\ y_n(t_1) & \dots & y_n(t_d) \end{pmatrix}$$

$i = 1..n$, $n \simeq 100$ à 1000 ; $x_{ij} \in \mathbb{R}$, $t \in [a, b]$.

Cadre d'étude

Code de calcul à *entrées vectorielles* et *sorties fonctionnelles*, **coûteux**.

$$\begin{pmatrix} x_{11} & \dots & x_{1p} \\ \vdots & \vdots & \vdots \\ x_{n1} & \dots & x_{np} \end{pmatrix} \longrightarrow \begin{pmatrix} y_1(t) \\ \vdots \\ y_n(t) \end{pmatrix} = \begin{pmatrix} y_1(t_1) & \dots & y_1(t_d) \\ \vdots & \vdots & \vdots \\ y_n(t_1) & \dots & y_n(t_d) \end{pmatrix}$$

$i = 1..n$, $n \simeq 100$ à 1000 ; $x_{ij} \in \mathbb{R}$, $t \in [a, b]$.

Objectif : prédiction de *données fonctionnelles* via un métamodèle.

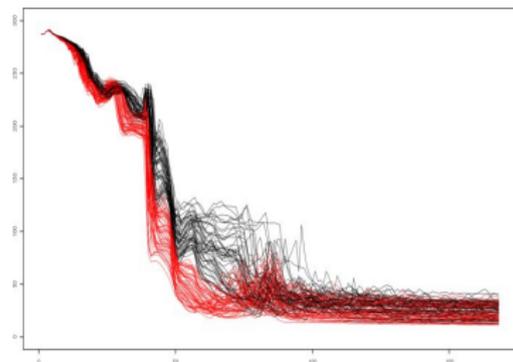


Fig.: 100 sorties du code Cathare.

Cadre d'étude

Code de calcul à *entrées vectorielles* et *sorties fonctionnelles*, **coûteux**.

$$\begin{pmatrix} x_{11} & \dots & x_{1p} \\ \vdots & \vdots & \vdots \\ x_{n1} & \dots & x_{np} \end{pmatrix} \longrightarrow \begin{pmatrix} y_1(t) \\ \vdots \\ y_n(t) \end{pmatrix} = \begin{pmatrix} y_1(t_1) & \dots & y_1(t_d) \\ \vdots & \vdots & \vdots \\ y_n(t_1) & \dots & y_n(t_d) \end{pmatrix}$$

$i = 1..n$, $n \simeq 100$ à 1000 ; $x_{ij} \in \mathbb{R}$, $t \in [a, b]$.

Objectif : prédiction de *données fonctionnelles* via un métamodèle.

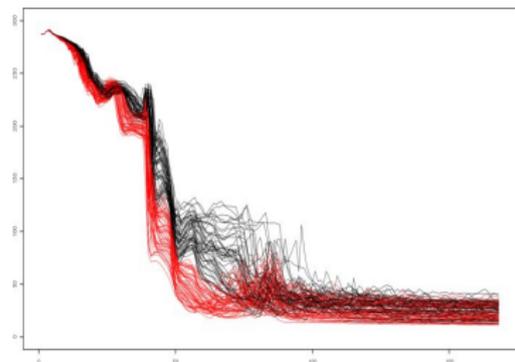


Fig.: 100 sorties du code Cathare.

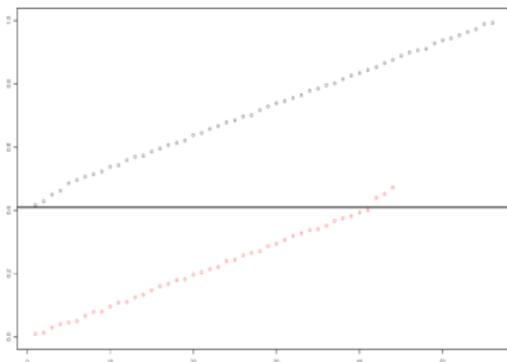


Fig.: Composante 4 des entrées.

But de la classification

→ Identifier des sous ensembles C_i des entrées / sorties tels que :

- pour $i \neq j$, C_i est bien distinct de C_j ;
- tous les éléments de C_i sont similaires.

But de la classification

→ Identifier des sous ensembles C_i des entrées / sorties tels que :

- pour $i \neq j$, C_i est bien distinct de C_j ;
- tous les éléments de C_i sont similaires.

Différents points de vue mènent à différentes solutions :

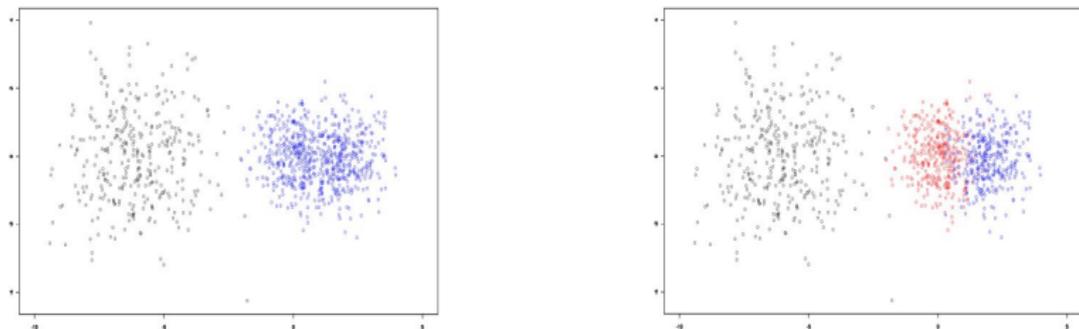


Fig.: Mélange de trois gaussiennes, $\mu = (-6, 0), (0, 0), (2, 0)$, $\sigma = 2, 0.5, 0.5$.

But de la classification

→ Identifier des sous ensembles C_i des entrées / sorties tels que :

- pour $i \neq j$, C_i est bien distinct de C_j ;
- tous les éléments de C_i sont similaires.

Différents points de vue mènent à différentes solutions :

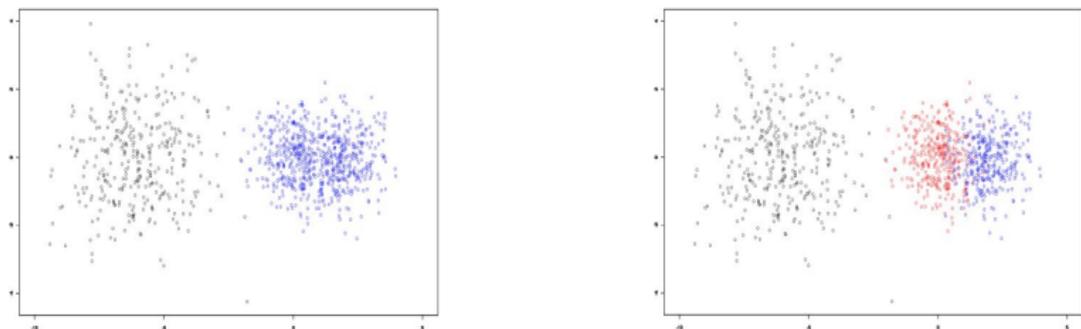


Fig.: Mélange de trois gaussiennes, $\mu = (-6, 0), (0, 0), (2, 0)$, $\sigma = 2, 0.5, 0.5$.

Critère objectif pour choisir le nombre de clusters K ?

Aperçu bibliographique

Principales méthodes de classification non supervisée :

- partitionnement autour de points principaux (*);
exemples : k -means, nuées dynamiques..;
- agglomérations hiérarchiques de "petits" clusters;
exemples : CAH, chameleon..;
- partitionnement dans un graphe représentant les données;
exemples : clustering spectral, b -coloriage..;
- clustering "évolutionnaire" d'inspiration biologique;
exemples : AntClust, CGA.. .

Aperçu bibliographique

Principales méthodes de classification non supervisée :

- partitionnement autour de points principaux (*);
exemples : k -means, nuées dynamiques..;
- agglomérations hiérarchiques de "petits" clusters;
exemples : CAH, chameleon..;
- partitionnement dans un graphe représentant les données;
exemples : clustering spectral, b -coloriage..;
- clustering "évolutionnaire" d'inspiration biologique;
exemples : AntClust, CGA.. .

Plan : introduction à (*) dans le cadre fonctionnel : $y(t) = \mu_k(t) + \tau_k(t)$,
puis analyse de la convergence des "fonctions principales".

Plan

- **A)** Partition des (entrées +) sorties en K clusters,
- apprentissage supervisé entrées – clusters.
- **B)** Réduction de la dimension des sorties,
- régression adaptée sur les coordonnées réduites.

Plan

- *A*) Partition des (entrées +) sorties en K clusters,
- apprentissage supervisé entrées – clusters.
- *B*) Réduction de la dimension des sorties,
- régression adaptée sur les coordonnées réduites.

1 Classification hiérarchique bayésienne

2 Clustering en dimension réduite

3 Tests

Description générale (C. Zhou, 2003)

Discrétisation : $y_{ij} = y_i(t_j)$, $i = 1..n$, $j = 1..D$.

- ① Quelle forme ont les courbes? Corrélations prises en compte via h :
 $y_{ij} = h(\beta_i, t_j) + e_{ij}$ avec $e_{ij} \sim \mathcal{N}(0, \sigma_e^2)$ erreur de mesure (..etc);

Exemple : $y_{ij} \simeq T\beta_i$, avec

$$T = \begin{pmatrix} 1 & t_1 & \dots & t_1^q \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_D & \dots & t_D^q \end{pmatrix}$$

Description générale (C. Zhou, 2003)

Discrétisation : $y_{ij} = y_i(t_j)$, $i = 1..n$, $j = 1..D$.

- 1 Quelle forme ont les courbes? Corrélations prises en compte via h :
 $y_{ij} = h(\beta_i, t_j) + e_{ij}$ avec $e_{ij} \sim \mathcal{N}(0, \sigma_e^2)$ erreur de mesure (..etc);

Exemple : $y_{ij} \simeq T\beta_i$, avec

$$T = \begin{pmatrix} 1 & t_1 & \dots & t_1^q \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_D & \dots & t_D^q \end{pmatrix}$$

- 2 Quelle loi suivent les paramètres des courbes?
 $\beta_i | z_i = k, \phi, K \sim \mathcal{g}(\beta_i, \phi_k)$ où z_i est l'étiquette de y_i ,
 ϕ_1, \dots, ϕ_K paramètres des clusters;

Exemple : $\phi_k = (\mu_k, C_k)$, $\mathcal{g}(\beta_i, \phi_k) \sim \mathcal{N}(\mu_k, C_k)$.

Description générale (C. Zhou, 2003)

Discrétisation : $y_{ij} = y_i(t_j)$, $i = 1..n$, $j = 1..D$.

- 1 Quelle forme ont les courbes? Corrélations prises en compte via h :
 $y_{ij} = h(\beta_i, t_j) + e_{ij}$ avec $e_{ij} \sim \mathcal{N}(0, \sigma_e^2)$ erreur de mesure (..etc);

Exemple : $y_{ij} \simeq T\beta_i$, avec

$$T = \begin{pmatrix} 1 & t_1 & \dots & t_1^q \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_D & \dots & t_D^q \end{pmatrix}$$

- 2 Quelle loi suivent les paramètres des courbes?
 $\beta_i | z_i = k, \phi, K \sim g(\beta_i, \phi_k)$ où z_i est l'étiquette de y_i ,
 ϕ_1, \dots, ϕ_K paramètres des clusters;

Exemple : $\phi_k = (\mu_k, C_k)$, $g(\beta_i, \phi_k) \sim \mathcal{N}(\mu_k, C_k)$.

- 3 la numérotation des clusters est interchangeable :
 $\mathbb{P}(z_1 = k_1, \dots, z_n = k_n | \pi, K) = \prod_{i=1}^n \pi_{k_i}$,
 $\pi_k =$ proportion du cluster k dans la population ;

Description générale - suite

- ④ Quelles lois suivent les macro-paramètres des clusters ?

$\phi_k \sim f(\phi_k | y, K)$ avec $y = (y_1, \dots, y_n)$;

Exemple : $\phi_k = (\mu_k, C_k)$ avec $\mu_k \sim \mathcal{N}(m, C)$,

$C_k^{-1} \sim \text{Wishart}(a, aC_0^{-1})$.

Description générale - suite

- ④ Quelles lois suivent les macro-paramètres des clusters ?

$\phi_k \sim f(\phi_k|y, K)$ avec $y = (y_1, \dots, y_n)$;

Exemple : $\phi_k = (\mu_k, C_k)$ avec $\mu_k \sim \mathcal{N}(m, C)$,

$$C_k^{-1} \sim \text{Wishart}(a, aC_0^{-1}).$$

- ⑤ Tout ne peut être déduit, donc au sommet : lois a priori ;

Exemples :

$$\pi \sim \text{Dirichlet}(\delta),$$

$$\sigma_e^{-2} \sim \text{Gamma}(a, b),$$

$$K \sim \text{Poisson}(\lambda) \dots$$

Description générale - suite

- 4 Quelles lois suivent les macro-paramètres des clusters ?

$\phi_k \sim f(\phi_k|y, K)$ avec $y = (y_1, \dots, y_n)$;

Exemple : $\phi_k = (\mu_k, C_k)$ avec $\mu_k \sim \mathcal{N}(m, C)$,

$C_k^{-1} \sim \text{Wishart}(a, aC_0^{-1})$.

- 5 Tout ne peut être déduit, donc au sommet : lois a priori ;

Exemples :

$\pi \sim \text{Dirichlet}(\delta)$,

$\sigma_e^{-2} \sim \text{Gamma}(a, b)$,

$K \sim \text{Poisson}(\lambda) \dots$

Puis ($\eta =$ paramètres globaux $\sigma_e^2, \delta, \dots$) :

$$p(y, z, \beta, \phi, \pi, K, \eta) = p(y|\beta, \eta)p(\beta|z, \phi, K, \eta) \quad [3]$$

$$p(z|\pi, K)p(\eta) \quad [2]$$

$$p(\phi|\eta, K)p(\pi|\eta, K)p(K) \quad [1]$$

Détermination de K

Idée : K dépend de la tolérance sur la représentativité des centroïdes μ_k , et donc des distributions des erreurs autour des centres.

Détermination de K

Idée : K dépend de la tolérance sur la représentativité des centroïdes μ_k , et donc des distributions des erreurs autour des centres.

..Si K augmente,

- interprétation / lisibilité du résultat diminue, mais
- précision (et complexité..) du modèle augmente.

Détermination de K

Idée : K dépend de la tolérance sur la représentativité des centroïdes μ_k , et donc des distributions des erreurs autour des centres.

..Si K augmente,

- interprétation / lisibilité du résultat diminue, mais
- précision (et complexité..) du modèle augmente.

Méthode

- 1 $K \leftarrow 1$;
- 2 tant que les clusters sont de taille suffisante (pour réduction de la dimension, classification / régression), et qu'on retrouve les groupes par classification supervisée + n -fold cross validation :
- 3 $K \leftarrow K + 1$, retour en 1.

Détermination de K

Idée : K dépend de la tolérance sur la représentativité des centroïdes μ_k , et donc des distributions des erreurs autour des centres.

..Si K augmente,

- interprétation / lisibilité du résultat diminue, mais
- précision (et complexité..) du modèle augmente.

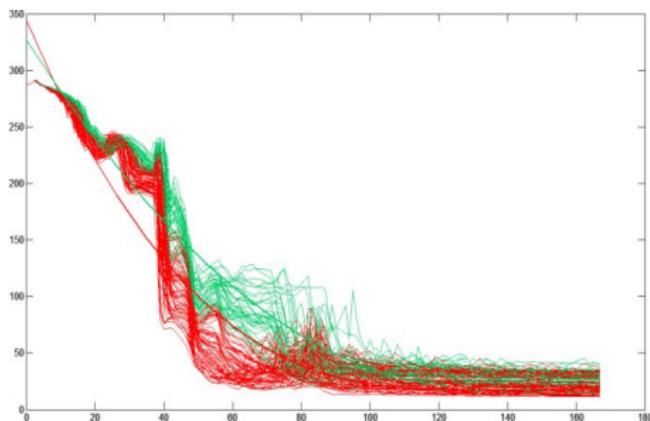
Méthode

- 1 $K \leftarrow 1$;
- 2 tant que les clusters sont de taille suffisante (pour réduction de la dimension, classification / régression), et qu'on retrouve les groupes par classification supervisée + n -fold cross validation :
- 3 $K \leftarrow K + 1$, retour en 1.

Options alternatives : "saut de distorsion maximal", comparaison à K clusters sur distribution uniforme ..etc.

Exemple : régression à effets aléatoires

Fig.: 2 clusters avec
CCT toolbox de S.
Gaffney (2003),
régressions linéaires :



Exemple : régression à effets aléatoires

Fig.: 2 clusters avec
CCT toolbox de S.
Gaffney (2003),
régressions linéaires :

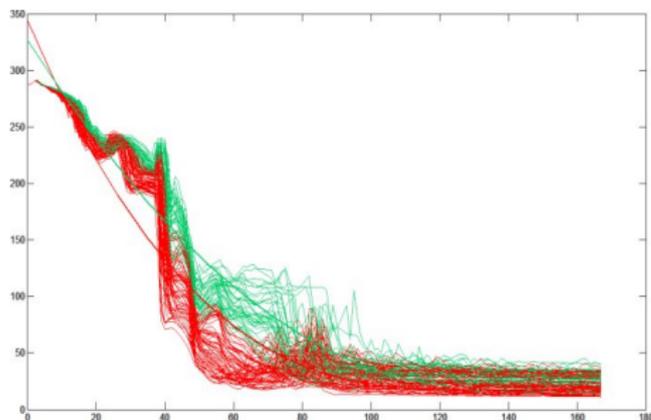
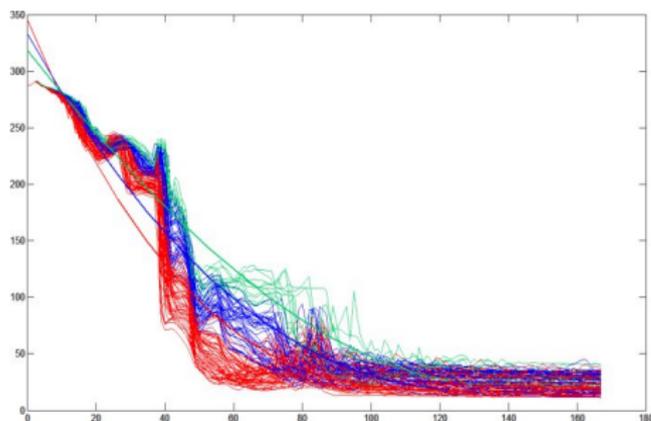


Fig.: ..3 clusters :



Bilan approches bayésiennes "fonctions principales"

Avantages

- + Ajout / suppression d'un y_i facile en "temps réel".
- + Traitement probabiliste des outliers.
- + **Manipulation mathématique simple.**

Bilan approches bayésiennes "fonctions principales"

Avantages

- + Ajout / suppression d'un y_i facile en "temps réel".
- + Traitement probabiliste des outliers.
- + **Manipulation mathématique simple.**

Inconvénients

- Estimation bayésienne de K : lourd, envisageable que si le nombre de points de discrétisation D est petit (t_1, \dots, t_D).
- **Sélection de modèle(s).**

Bilan approches bayésiennes "fonctions principales"

Avantages

- + Ajout / suppression d'un y_i facile en "temps réel".
- + Traitement probabiliste des outliers.
- + **Manipulation mathématique simple.**

Inconvénients

- Estimation bayésienne de K : lourd, envisageable que si le nombre de points de discrétisation D est petit (t_1, \dots, t_D).
- **Sélection de modèle(s).**

Autre vision possible : on n'impose pas de modèle paramétrique sur les clusters, qui sont alors déterminés par les relations entre les courbes.

Plan

1 Classification hiérarchique bayésienne

2 Clustering en dimension réduite

3 Tests

Remarques pratiques

Une fonction n'est jamais manipulée comme un objet de dimension infinie :
 $f \in \mathcal{F}([a, b]) = (f(t_1), \dots, f(t_D)) \in \mathbb{R}^D$ en informatique.

Remarques pratiques

Une fonction n'est jamais manipulée comme un objet de dimension infinie :
 $f \in \mathcal{F}([a, b]) = (f(t_1), \dots, f(t_D)) \in \mathbb{R}^D$ en informatique.

Hypothèse : les t_j sont assez rapprochés pour calculer les intégrales
 $\int y_i(t)b(t)dt$ avec une erreur négligeable, $b \in L_2([a, b])$.

$\Rightarrow D$ trop grand pour considérer $(y_i(t_1), \dots, y_i(t_D))$ comme une bonne représentation (facile à interpréter ..etc) de y_i .

" $D = +\infty$ " \rightarrow passage en dimension finie.

Remarques pratiques

Une fonction n'est jamais manipulée comme un objet de dimension infinie :
 $f \in \mathcal{F}([a, b]) = (f(t_1), \dots, f(t_D)) \in \mathbb{R}^D$ en informatique.

Hypothèse : les t_j sont assez rapprochés pour calculer les intégrales
 $\int y_i(t)b(t)dt$ avec une erreur négligeable, $b \in L_2([a, b])$.

$\Rightarrow D$ trop grand pour considérer $(y_i(t_1), \dots, y_i(t_D))$ comme une bonne représentation (facile à interpréter ..etc) de y_i .

" $D = +\infty$ " \rightarrow passage en dimension finie.

Décomposition sur une base

On peut obtenir un nombre arbitraire de coefficients de décomposition sur une base hilbertienne $(b_j)_{j \in \mathbb{N}}$ de $L_2([a, b])$.

..Mais pas un nombre infini.

\Rightarrow Les coefficients $c_{ij} = \langle y_i, b_j \rangle$ doivent converger rapidement vers 0.
On ne conserve que les premiers pour représenter y_i .

Hypothèse principale

Base orthonormée $b = b_0, \dots, b_j, \dots$ de $L_2([a, b])$ fixée,
 n courbes $y_i : [a, b] \rightarrow \mathbb{R}$, $c_{ij} = \int y_i(t)b_j(t)dt$.

Hypothèse

$\exists(\varphi_j)_{j \in \mathbb{N}}$ suite croissante de \mathbb{R}^{+*} avec $\varphi_j \rightarrow +\infty$, $\exists R > 0$ tels que :

$$\forall i \in [1, n], \quad \sum_{j=0}^{+\infty} \varphi_j c_{ij}^2 < R^2 .$$

E = espace des fonctions vérifiant cette propriété.

Hypothèse principale

Base orthonormée $b = b_0, \dots, b_j, \dots$ de $L_2([a, b])$ fixée,
 n courbes $y_i : [a, b] \rightarrow \mathbb{R}$, $c_{ij} = \int y_i(t)b_j(t)dt$.

Hypothèse

$\exists(\varphi_j)_{j \in \mathbb{N}}$ suite croissante de \mathbb{R}^{+*} avec $\varphi_j \rightarrow +\infty$, $\exists R > 0$ tels que :

$$\forall i \in [1, n], \quad \sum_{j=0}^{+\infty} \varphi_j c_{ij}^2 < R^2 .$$

E = espace des fonctions vérifiant cette propriété.

\Rightarrow Les courbes $y \in E$ ne sont pas trop "dispersées" :

Lemme

$$\forall y \in E, \quad \|y\|_{L_2} \leq \frac{R^2}{\varphi_0} = B .$$

Définitions, objectifs

Courbes y_i classées suivant leurs distances aux centres c_j , $j = 1..K$.

Π_d = opérateur de projection sur les d premières fonctions de base.

Distorsion empirique en dimension d (resp. ∞ sans $\{d, \Pi_d\}$) :

$$W_{nd}(c) = \frac{1}{n} \sum_{i=1}^n \min_{j=1..K} \|\Pi_d(y_i) - \Pi_d(c_j)\|^2$$

Distorsion moyenne en dimension d (resp. ∞ sans $\{d, \Pi_d\}$) :

$$W_d(c) = \mathbb{E}_y \min_{j=1..K} \|\Pi_d(y) - \Pi_d(c_j)\|^2$$

But : trouver un jeu de centroïdes $c^* \in E^K$ minimisant $W(c)$.

Définitions, objectifs

Courbes y_i classées suivant leurs distances aux centres c_j , $j = 1..K$.

$\Pi_d =$ opérateur de projection sur les d premières fonctions de base.

Distorsion empirique en dimension d (resp. ∞ sans $\{d, \Pi_d\}$) :

$$W_{nd}(c) = \frac{1}{n} \sum_{i=1}^n \min_{j=1..K} \|\Pi_d(y_i) - \Pi_d(c_j)\|^2$$

Distorsion moyenne en dimension d (resp. ∞ sans $\{d, \Pi_d\}$) :

$$W_d(c) = \mathbb{E}_y \min_{j=1..K} \|\Pi_d(y) - \Pi_d(c_j)\|^2$$

But : trouver un jeu de centroïdes $c^* \in E^K$ minimisant $W(c)$.

En pratique, supposant avoir c_{nd}^* minimiseur global de W_{nd} , quelle erreur effectue-t-on en utilisant ces centres en dimension infinie ?

Question principale

$$\lim_{n,d \rightarrow +\infty} W(c_{nd}^*) - W(c^*) = 0 \quad ?$$

Résultat

Proposition 1 : majoration de l'erreur de projection ;

$$\sup_{c \in E^k} \{W(c) - W_d(c)\} \leq \frac{2R^2}{\varphi_{d+1}} .$$

Proposition 2 : pour tout $\alpha > 0$, $W_d(c_{nd}^*) - W_d(c_d^*) \leq \alpha$ avec probabilité au moins $P(\alpha)$ (T. Linder, G. Lugosi, K. Zeger, 1994) :

$$P(\alpha) = \max \left(0, 1 - 4(2n)^{k(d+1)} e^{\frac{-n\alpha^2}{512B^2}} \right) .$$

Résultat

Proposition 1 : majoration de l'erreur de projection ;

$$\sup_{c \in E^k} \{W(c) - W_d(c)\} \leq \frac{2R^2}{\varphi_{d+1}}.$$

Proposition 2 : pour tout $\alpha > 0$, $W_d(c_{nd}^*) - W_d(c_d^*) \leq \alpha$ avec probabilité au moins $P(\alpha)$ (T. Linder, G. Lugosi, K. Zeger, 1994) :

$$P(\alpha) = \max \left(0, 1 - 4(2n)^{k(d+1)} e^{\frac{-n\alpha^2}{512B^2}} \right).$$

Théorème :

Soit $\alpha > 0$, et $d \in \mathbb{N}^*$. Alors avec probabilité au moins $P(\alpha)$:

$$W(c_{nd}^*) - W(c^*) \leq \alpha + \frac{4R^2}{\varphi_{d+1}}.$$

Bilan, ouvertures

Alternative : méthode des *projections aléatoires* (G. Biau, L. Devroye, G. Lugosi, 2005) :

$c_{i1}, \dots, c_{ij}, \dots$ représentation orthonormée de y_i ,

- 1 pour N_{kl} variables aléatoires gaussiennes, calculer $c_i^{(k)} = \sum_{j=1}^{+\infty} N_{kj} c_{ij}$;
- 2 clustering des données $c_i^{(1)}, \dots, c_i^{(d)}$ dans l'espace \mathbb{R}^d .

Posant alors $d = \lceil 4(\varepsilon^2/2 - \varepsilon^3/3)^{-1} \log(n/\sqrt{\delta}) \rceil$,

avec une probabilité $\geq (1 - \delta)(1 - e^{-x})$:

$$W(c_{nd}^*) - W(c^*) \leq \frac{24kB + 12B\sqrt{2x}}{\sqrt{n}} + 4\varepsilon B.$$

Bilan, ouvertures

Alternative : méthode des *projections aléatoires* (G. Biau, L. Devroye, G. Lugosi, 2005) :

$c_{i1}, \dots, c_{ij}, \dots$ représentation orthonormée de y_i ,

- 1 pour N_{kl} variables aléatoires gaussiennes, calculer $c_i^{(k)} = \sum_{j=1}^{+\infty} N_{kj} c_{ij}$;
- 2 clustering des données $c_i^{(1)}, \dots, c_i^{(d)}$ dans l'espace \mathbb{R}^d .

Posant alors $d = \lceil 4(\varepsilon^2/2 - \varepsilon^3/3)^{-1} \log(n/\sqrt{\delta}) \rceil$,

avec une probabilité $\geq (1 - \delta)(1 - e^{-x})$:

$$W(c_{nd}^*) - W(c^*) \leq \frac{24kB + 12B\sqrt{2x}}{\sqrt{n}} + 4\varepsilon B.$$

Problème

En pratique, il faut aussi une hypothèse de décroissance forte des c_{ij} .

Autres pistes : bases locales pdt. le clustering, courbes principales ...

Plan

- 1 Classification hiérarchique bayésienne
- 2 Clustering en dimension réduite
- 3 Tests

Comparaisons I

Méthode	Taux d'erreur
K-Means	37%
Hierarchique + Ward	2%
Clustering Spectral	0%
Bayésien RL	12%

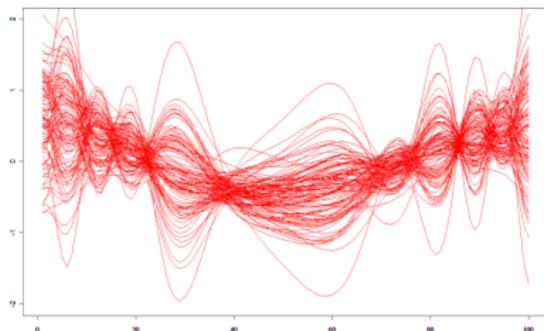
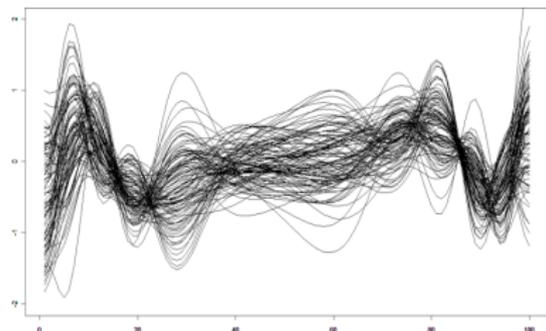
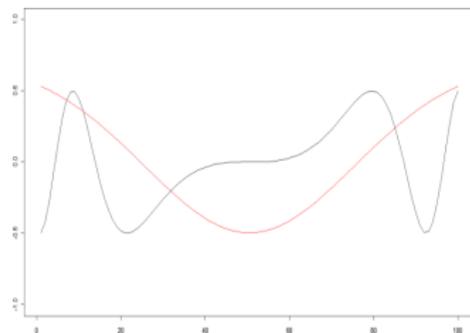


Fig.: $f_{a,b,c}$ sur $[-4, 4]$, groupe 1 à gauche, **groupe 2** à droite.

Fonction $f_{a,b,c} : x \mapsto \mu_k(x) + a \sin(x) + b \cos(x^2) + c \frac{x^2 - x^3}{40}$,

avec $\mu_1(x) = 0.5 \sin((0.5x)^3)$, $\mu_2(x) = -3e^{-t^2/8} / \sqrt{2\pi} + 0.7$,

(a, b, c) perturbation aléatoire gaussienne.

Comparaisons II

Méthode	Taux d'erreur
K-Means	47%
Hierarchique + Ward	46%
Clustering Spectral	46%
Bayésien RL	56%

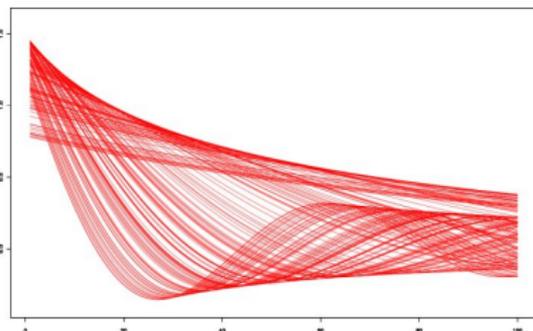
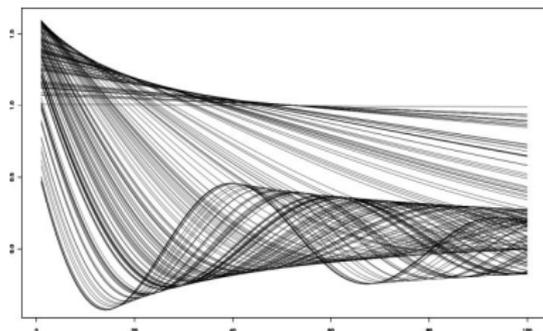
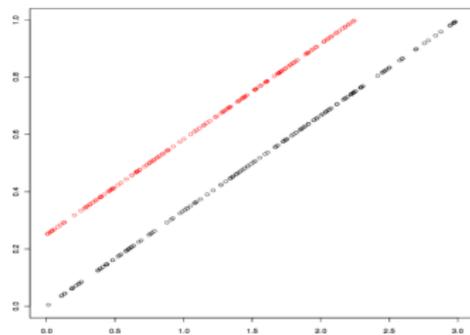


Fig.: $f_{a,b}$ sur $[1, 5]$, groupe 1 à gauche, groupe 2 à droite.

Fonction $f_{a,b} : x \mapsto \frac{\sin(ax)}{x} + e^{-bx}$.

(a, b) uniforme sur le segment $3b = a$, $b \in [0, 1]$ pour 1 à 150,
puis uniforme sur $3(b - 0.25) = a$, $b \in [0.25, 1]$ (courbes 151 à 300).

Conclusion

Model-based clustering

Les données suivent un certain modèle choisi dans un "dictionnaire" ;
 z_i déterminé par maximisation de $\mathbb{P}(z_i | \dots)$.

Model-free clustering

Données organisées en groupes de similarité de nature inconnue ;
Clustering sur similarités $s(y_i, y_j)$ supposées connues.

Conclusion

Model-based clustering

Les données suivent un certain modèle choisi dans un "dictionnaire" ;
 z_i déterminé par maximisation de $\mathbb{P}(z_i | \dots)$.

Model-free clustering

Données organisées en groupes de similarité de nature inconnue ;
Clustering sur similarités $s(y_i, y_j)$ supposées connues.

Classification suivant un modèle :

- interprétation, analyse et manipulation plus faciles ;
- clusters peu fiables si aucune connaissances a priori.

⇒ à utiliser en priorité dans les cas où la forme des courbes est connue.

Conclusion

Model-based clustering

Les données suivent un certain modèle choisi dans un "dictionnaire" ; z_i déterminé par maximisation de $\mathbb{P}(z_i | \dots)$.

Model-free clustering

Données organisées en groupes de similarité de nature inconnue ; Clustering sur similarités $s(y_i, y_j)$ supposées connues.

Classification suivant un modèle :

- interprétation, analyse et manipulation plus faciles ;
- clusters peu fiables si aucune connaissances a priori.

⇒ à utiliser en priorité dans les cas où la forme des courbes est connue.

"Best of two worlds" :

- détermination automatique de modèle, et/ou
- similarité vectorielle sans pertes d'information.