Modèles de mélange pour la régression en grande dimension, application aux données fonctionnelles.

Emilie Devijver

Université Paris-Sud XI
Select-INRIA

Observations : $((x_1, y_1), \ldots, (x_n, y_n)) \in (\mathbb{R}^p \times \mathbb{R}^q)^n$, où $p \times q >> n \times q$

But : trouver une structure pour Y sachant $X = x$.

Observations : $((x_1, y_1), \ldots, (x_n, y_n)) \in (\mathbb{R}^p \times \mathbb{R}^q)^n$, où $p \times q >> n \times q$

But : trouver une structure pour Y sachant $X = x$.

Notre approche :

- Estimer le nombre de composantes,
- bien estimer le modèle dans chaque composante,
- identifier les variables explicatives,
- identifier les groupes associés aux observations.

Motivation 1

- x : consommation individuelle pour le jour d
- y : consommation individuelle pour le jour $d+1$

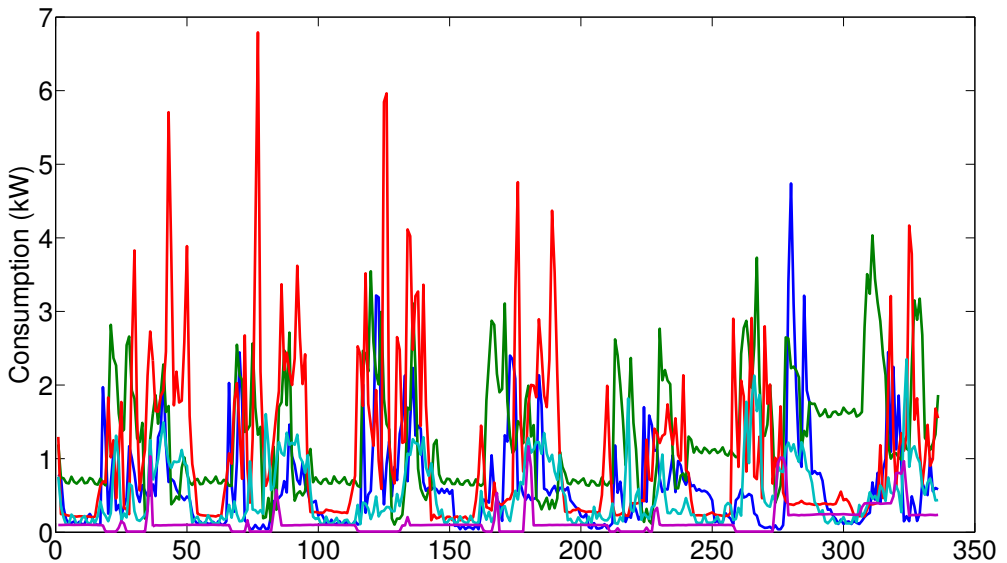


Figure : Échantillon de 5 consommateurs sur une semaine en hiver.

Motivation 2

- Modèles de mélange en régression *Städler et al., 2010*
- Sélection de modèles pour les mélanges en grande dimension *Meynet, 2012*
- Sélection de modèles pour les densités conditionnelles *Cohen, Le Pennec, 2011*
- Estimation par faible rang *Giraud, 2011*

Vue d'ensemble des résultats

- Deux procédures pour la classification non supervisée en grande dimension en régression
- Une inégalité oracle ℓ_1 satisfaite par le Lasso dans le cadre de modèles de mélanges en régression
- Une inégalité oracle ℓ_0 satisfaite par chaque procédure pour la sélection de modèles
- Application d'une procédure sur des données réelles électriques

Vue d'ensemble des résultats

- *Deux procédures pour la classification non supervisée en grande dimension en régression*
- Une inégalité oracle ℓ_1 satisfaite par le Lasso dans le cadre de modèles de mélanges en régression
- *Une inégalité oracle ℓ_0 satisfaite par chaque procédure pour la sélection de modèles*
- *Application d'une procédure sur des données réelles électriques*

Si y_i, connaissant x_i, appartient à la classe k, il existe β_k et Σ_k tels que

$$y_i = \beta_k x_i + \varepsilon_i,$$

où $\varepsilon_i \sim \mathcal{N}_q(0, \Sigma_k)$.

Si y_i, connaissant x_i, appartient à la classe k, il existe β_k et Σ_k tels que

$$y_i = \beta_k x_i + \varepsilon_i,$$

où $\varepsilon_i \sim \mathcal{N}_q(0, \Sigma_k)$.

- Les variables $Y_i|X_i$ sont indépendantes, pour tout $i = 1, \ldots, n$;
- les variables $Y_i|X_i = x_i \sim f_\xi(y|x_i)dy$, avec

$$f_\xi(y|x_i) = \sum_{k=1}^{K} \frac{\pi_k}{(2\pi)^{q/2}\det(\Sigma_k)^{1/2}} \exp\left(-\frac{(y-\beta_k x)^t \Sigma_k^{-1}(y-\beta_k x)}{2}\right)$$

$$\xi = (\pi_1, \ldots, \pi_K, \beta_1, \ldots, \beta_K, \Sigma_1, \ldots, \Sigma_K) \in \left(\Pi_K \times (\mathbb{R}^{q\times p})^k \times (\mathbb{S}_+^q)^k\right)$$

Si y_i, connaissant x_i, appartient à la classe k, il existe β_k et Σ_k tels que

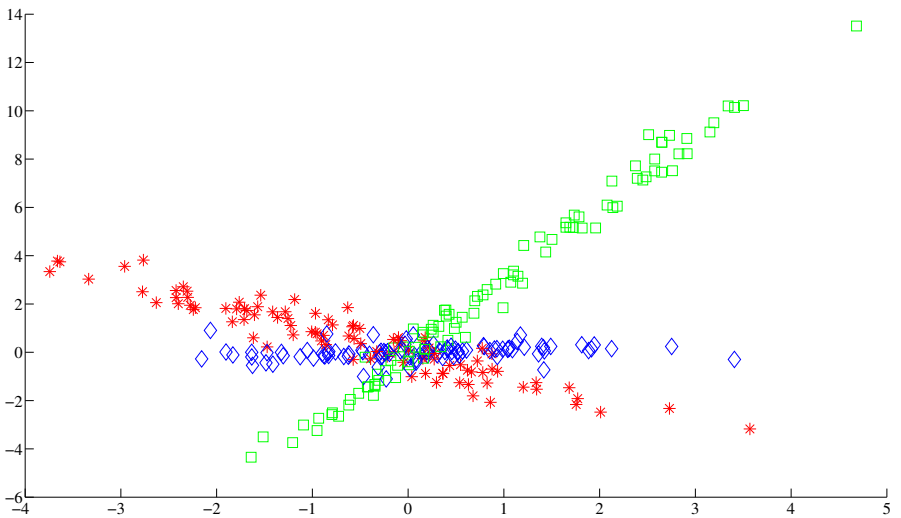
$$y_i = \beta_k x_i + \varepsilon_i,$$

où $\varepsilon_i \sim \mathcal{N}_q(0, \Sigma_k)$.

- Les variables $Y_i|X_i$ sont indépendantes, pour tout $i = 1, \ldots, n$;
- les variables $Y_i|X_i = x_i \sim f_\xi(y|x_i)dy$, avec

$$f_\xi(y|x_i) = \sum_{k=1}^{K} \frac{\pi_k}{(2\pi)^{q/2}\det(\Sigma_k)^{1/2}} \exp\left(-\frac{(y-\beta_k x)^t \Sigma_k^{-1}(y-\beta_k x)}{2}\right)$$

$$\xi = (\pi_1, \ldots, \pi_K, \beta_1, \ldots, \beta_K, \Sigma_1, \ldots, \Sigma_K) \in \left(\Pi_K \times (\mathbb{R}^{q\times p})^k \times (\mathbb{S}_+^q)^k\right)$$



Pour avoir une estimation **invariante par changement d'échelle** et une optimisation **convexe**, on reparamétrise.

$$
\begin{aligned}
{}^t\tilde{P}_k\tilde{P}_k &= \Sigma_k \\
P_k &= \tilde{P}_k^{-1} \\
\Phi_k &= P_k\beta_k \\
\theta &= (\pi_1,\ldots,\pi_K,\Phi_1,\ldots,\Phi_K,P_1,\ldots,P_K) \in \Theta_K \\
\Theta_K &= \left(\Pi_K \times (\mathbb{R}^{q\times p})^K \times T_q^K\right)
\end{aligned}
$$

pour tout $k \in \{1,\ldots,K\}$.

Pour avoir une estimation **invariante par changement d'échelle** et une optimisation **convexe**, on reparamétrise.

$$
\begin{aligned}
{}^t\tilde{P}_k\tilde{P}_k &= \Sigma_k \\
P_k &= \tilde{P}_k^{-1} \\
\Phi_k &= P_k\beta_k \\
\theta &= (\pi_1,\ldots,\pi_K,\Phi_1,\ldots,\Phi_K,P_1,\ldots,P_K) \in \Theta_K \\
\Theta_K &= \left(\Pi_K \times (\mathbb{R}^{q\times p})^K \times T_q^K\right)
\end{aligned}
$$

pour tout $k \in \{1,\ldots,K\}$.

Réduction de la dimension $\rightsquigarrow \Theta_K = \left(\Pi_K \times (\mathbb{R}^{q\times p})^K \times (\mathbb{R}^q)^K\right)$

La logvraisemblance est alors

$$\ell(\theta, \mathbf{y}, \mathbf{x}) = \sum_{i=1}^{n} \log \left(\sum_{k=1}^{K} \frac{\pi_k \det(P_k)}{(2\pi)^{q/2}} \exp \left(- \frac{(P_k y_i - x_i \Phi_k)^t (P_k y_i - x_i \Phi_k)}{2} \right) \right)$$

On définit le maximum de vraisemblance par

$$\hat{\theta}_0 := \underset{\theta \in \Theta_k}{\operatorname{argmin}} \left\{ -\frac{1}{n} \ell(\theta, \mathbf{y}, \mathbf{x}) \right\}$$

La logvraisemblance est alors

$$\ell(\theta, \mathbf{y}, \mathbf{x}) = \sum_{i=1}^{n} \log \left(\sum_{k=1}^{K} \frac{\pi_k \det(P_k)}{(2\pi)^{q/2}} \exp \left(-\frac{(P_k y_i - x_i \Phi_k)^t (P_k y_i - x_i \Phi_k)}{2} \right) \right)$$

On définit le maximum de vraisemblance par

$$\hat{\theta}_0 := \underset{\theta \in \Theta_k}{\operatorname{argmin}} \left\{ -\frac{1}{n} \ell(\theta, \mathbf{y}, \mathbf{x}) \right\}$$

et l'estimateur du Lasso par

$$\hat{\theta}_\lambda := \underset{\theta \in \Theta_K}{\operatorname{argmin}} \left\{ -\frac{1}{n} \ell_\lambda(\theta, \mathbf{y}, \mathbf{x}) \right\}$$

où

$$\ell_\lambda(\theta, \mathbf{y}, \mathbf{x}) = \ell(\theta, \mathbf{y}, \mathbf{x}) - n\lambda \sum_{k=1}^{K} \pi_k ||\Phi_k||_1$$

avec $||\Phi_k||_1 = \sum_{j=1}^{p} \sum_{z=1}^{q} |[\Phi_k]_{z,j}|$, avec $\lambda \geq 0$ à spécifier.

Maximisation de la logvraisemblance (pénalisée ou non) d'une densité de mélange
$\rightarrow$ algorithme EM (*Dempster et al.*)

- **Étape E** : calcul des affectations
- **Étape M** : calcul des paramètres qui maximisent la vraisemblance

Maximisation de la logvraisemblance (pénalisée ou non) d'une densité de mélange
$\rightarrow$ algorithme EM (*Dempster et al.*)

- **Étape E** : calcul des affectations
- **Étape M** : calcul des paramètres qui maximisent la vraisemblance

Initialisation

Plusieurs fois :

- On initialise
  - initialisation des affectations : k-means sur les couples
  - initialisation des paramètres : modèle linéaire dans chaque classe
- On fait un peu tourner l'algorithme

On garde l'initialisation qui majore la vraisemblance après quelques itérations

Maximisation de la logvraisemblance (pénalisée ou non) d'une densité de mélange
$\rightarrow$ algorithme EM (*Dempster et al.*)

- **Étape E** : calcul des affectations
- **Étape M** : calcul des paramètres qui maximisent la vraisemblance

Initialisation

Plusieurs fois :

- On initialise
  - initialisation des affectations : k-means sur les couples
  - initialisation des paramètres : modèle linéaire dans chaque classe
- On fait un peu tourner l'algorithme

On garde l'initialisation qui majore la vraisemblance après quelques itérations

Conditions d'arrêt

- convergence relative des paramètres
- convergence relative de la logvraisemblance
- nombre maximum d'itérations

Soit $K \in \mathcal{K}$ le nombre de composantes.

Soit $K \in \mathcal{K}$ le nombre de composantes.

Definition

On dit que les variables indicées par $(j_1, j_2) \in J^c$ *sont inactives pour le clustering si* $[\phi_1]_{j_1,j_2} = \ldots = [\phi_k]_{j_1,j_2} = 0$.

Soit $K \in \mathcal{K}$ le nombre de composantes.

Definition

On dit que les variables indicées par $(j_1, j_2) \in J^c$ *sont inactives pour le clustering si* $[\phi_1]_{j_1,j_2} = \ldots = [\phi_k]_{j_1,j_2} = 0$.

Quel choix pour λ ?

Soit $K \in \mathcal{K}$ le nombre de composantes.

Definition

On dit que les variables indicées par $(j_1, j_2) \in J^c$ *sont inactives pour le clustering si* $[\phi_1]_{j_1,j_2} = \ldots = [\phi_k]_{j_1,j_2} = 0$.

Quel choix pour λ **?** $\rightsquigarrow$ grille de paramètres de régularisation G_K explicite d'après les formules de mise à jour de l'algorithme EM.

Soit $K \in \mathcal{K}$ le nombre de composantes.

Definition

On dit que les variables indicées par $(j_1, j_2) \in J^c$ *sont inactives pour le clustering si* $[\phi_1]_{j_1,j_2} = \ldots = [\phi_k]_{j_1,j_2} = 0$.

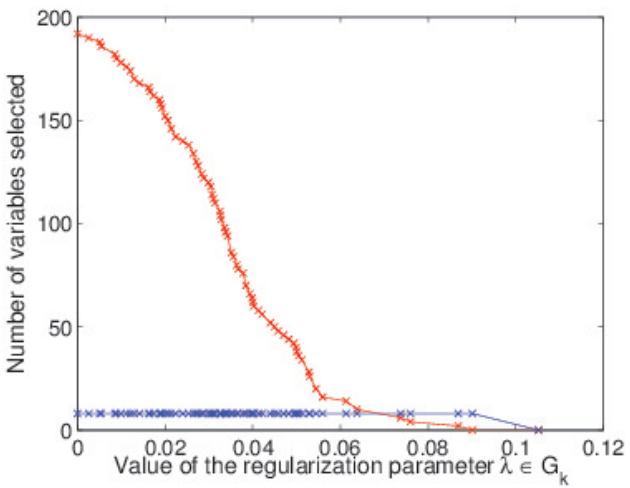
Quel choix pour λ **?** $\rightsquigarrow$ grille de paramètres de régularisation G_K explicite d'après les formules de mise à jour de l'algorithme EM.
Pour tout $\lambda \in G_K$, approximation de l'estimateur du Lasso.

Soit $K \in \mathcal{K}$ le nombre de composantes.

Definition

On dit que les variables indicées par $(j_1, j_2) \in J^c$ *sont inactives pour le clustering si* $[\phi_1]_{j_1,j_2} = \ldots = [\phi_k]_{j_1,j_2} = 0$.

Quel choix pour λ **?** $\rightsquigarrow$ grille de paramètres de régularisation G_K explicite d'après les formules de mise à jour de l'algorithme EM.
Pour tout $\lambda \in G_K$, approximation de l'estimateur du Lasso.
On en déduit l'ensemble des variables actives, pour tout $\lambda \in G_K$, pour tout $K \in \mathcal{K}$.

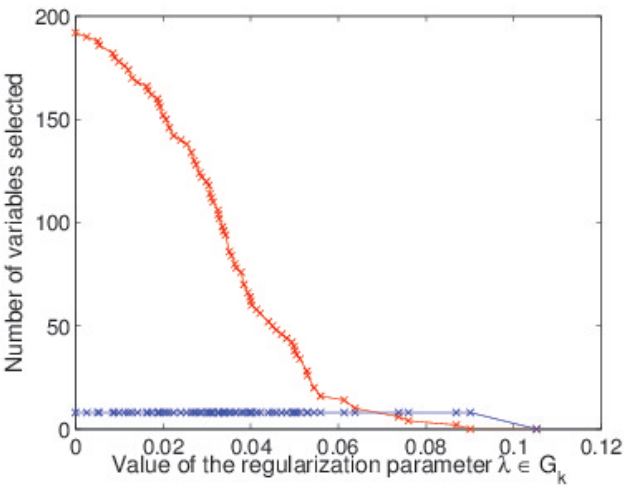


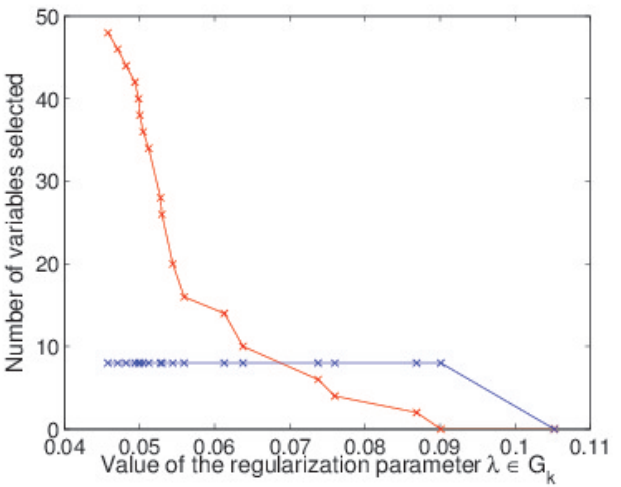
Soit $K \in \mathcal{K}$ le nombre de composantes.

Definition

On dit que les variables indicées par $(j_1, j_2) \in J^c$ *sont inactives pour le clustering si* $[\phi_1]_{j_1,j_2} = \ldots = [\phi_k]_{j_1,j_2} = 0$.

Quel choix pour λ **?** $\rightsquigarrow$ grille de paramètres de régularisation G_K explicite d'après les formules de mise à jour de l'algorithme EM.
Pour tout $\lambda \in G_K$, approximation de l'estimateur du Lasso.
On en déduit l'ensemble des variables actives, pour tout $\lambda \in G_K$, pour tout $K \in \mathcal{K}$.





Soit $K \in \mathcal{K}$, $\lambda \in G_K$.

Procédure Lasso-MLE

On réestime nos paramètres par l'**estimateur du maximum de vraisemblance** restreint aux variables actives.

$$\mathcal{S}_{(K,J)} = \left\{ y \in \mathbb{R}^q \mapsto s_\xi^{(K,J)}(y|x) \right\}$$

$$\Xi_{(K,J)} = \Pi_K \times (\mathbb{R}^J)^K \times (\mathbb{S}_q^{++})^K$$

Soit $K \in \mathcal{K}$, $\lambda \in G_K$.

Procédure Lasso-MLE

On réestime nos paramètres par l'**estimateur du maximum de vraisemblance** restreint aux variables actives.

$$\mathcal{S}_{(K,J)} = \left\{ y \in \mathbb{R}^q \mapsto s_\xi^{(K,J)}(y|x) \right\}$$
$$\Xi_{(K,J)} = \Pi_K \times (\mathbb{R}^J)^K \times (\mathbb{S}_q^{++})^K$$

Procédure Lasso-Rang

On réestime nos paramètres par l'**estimateur du maximum de vraisemblance de faible rang** restreint aux variables actives.

$$\mathcal{S}_{(K,J,R)} = \left\{ y \in \mathbb{R}^q \mapsto s_\xi^{(K,J,R)}(y|x) \right\}$$
$$\check{\Xi}_{(K,J,R)} = \Pi_K \times \Psi_{(K,J,R)} \times (\mathbb{S}_q^{++})^K$$

D'un point de vue pratique

Heuristique des pentes

$$\hat{m} = \underset{m \in \mathcal{M}}{\operatorname{argmin}} \left\{ -L(\hat{\theta}_m) + \frac{pen(m)}{n} \right\}$$

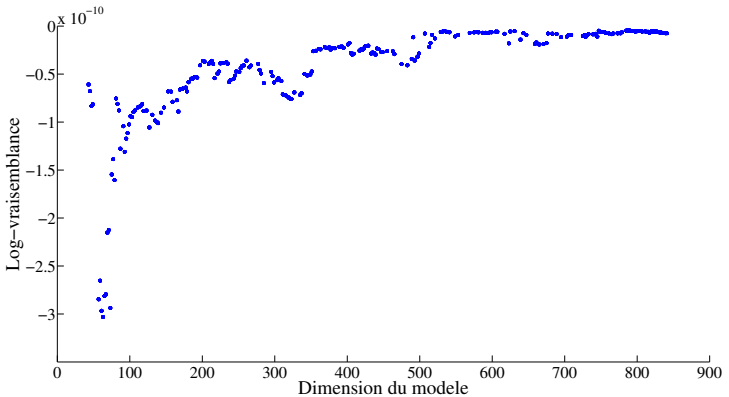
$$pen(m) = \kappa D_m$$

D'un point de vue pratique

Heuristique des pentes

$$\hat{m} = \underset{m \in \mathcal{M}}{\operatorname{argmin}} \left\{ -L(\hat{\theta}_m) + \frac{pen(m)}{n} \right\}$$

$$pen(m) = \kappa D_m$$

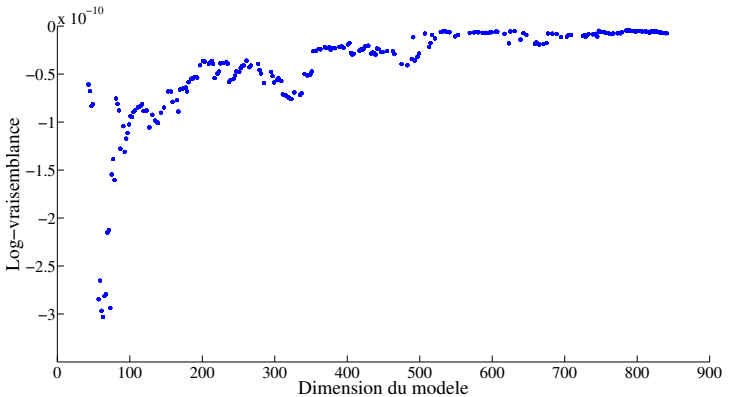


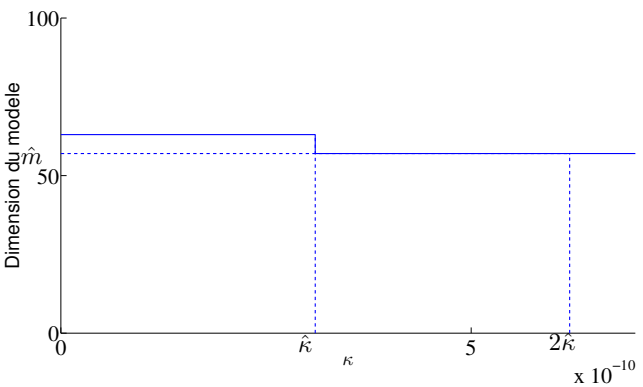
D'un point de vue pratique

Heuristique des pentes

$$\hat{m} = \underset{m \in \mathcal{M}}{\operatorname{argmin}} \left\{ -L(\hat{\theta}_m) + \frac{pen(m)}{n} \right\}$$

$$pen(m) = \kappa D_m$$





D'un point de vue théorique *Massart, Cohen, Le Pennec*

- $\mathcal{S} = (S_m)_{m \in \mathcal{M}}$ la collection de densités conditionnelles,
- $\tilde{\mathcal{S}} = (S_m)_{m \in \tilde{\mathcal{M}}}$ une sous-collection aléatoire de $\mathcal{S}$,
- $(\hat{s}_m)_{m \in \tilde{\mathcal{M}}}$ la collection des η-maxima de vraisemblance dans $\tilde{\mathcal{S}}$.

Si la collection de modèle vérifie les hypothèses K et B_τ, et pour tous les modèles m, les hypothèses H_m et Sep_m,

D'un point de vue théorique *Massart, Cohen, Le Pennec*

- $\mathcal{S} = (S_m)_{m\in\mathcal{M}}$ la collection de densités conditionnelles,
- $\tilde{\mathcal{S}} = (S_m)_{m\in\tilde{\mathcal{M}}}$ une sous-collection aléatoire de $\mathcal{S}$,
- $(\hat{s}_m)_{m\in\tilde{\mathcal{M}}}$ la collection des η-maxima de vraisemblance dans $\tilde{\mathcal{S}}$.

Si la collection de modèle vérifie les hypothèses K et B_τ, et pour tous les modèles m, les hypothèses H_m et Sep_m,

Alors, pour tout $\rho \in (0,1)$ et $C_1 > 1$, il existe κ_0 et C_2 tels que pour tout $\kappa > \kappa_0$, si

$$pen(m) \geq \frac{\kappa}{n}(\mathcal{D}_m + (1\vee\tau)x_m)$$

alors $\hat{m} = \underset{m\in\tilde{\mathcal{M}}}{\operatorname{argmin}} \left(\gamma_n(\hat{s}_m) + pen(m) + \eta'\right)$ vérifie

$$E(JKL_\rho^{\otimes n}(s^*, \hat{s}_{\hat{m}})) \leq C_1 E\left(\inf_{m\in\tilde{\mathcal{M}}}\left(\inf_{t\in S_m} KL^{\otimes n}(s^*, t)\right) + 2pen(m)\right)$$
$$+ C_2(1\vee\tau)\frac{\Omega^2}{n} + \frac{\eta'+\eta}{n}.$$

D'un point de vue théorique *Massart, Cohen, Le Pennec*

- $\mathcal{S} = (S_m)_{m\in\mathcal{M}}$ la collection de densités conditionnelles,
- $\tilde{\mathcal{S}} = (S_m)_{m\in\tilde{\mathcal{M}}}$ une sous-collection aléatoire de $\mathcal{S}$,
- $(\hat{s}_m)_{m\in\tilde{\mathcal{M}}}$ la collection des η-maxima de vraisemblance dans $\tilde{\mathcal{S}}$.

Si la collection de modèle vérifie les hypothèses K et B_τ, et pour tous les modèles m, les hypothèses H_m et Sep_m,

Alors, pour tout $\rho \in (0,1)$ et $C_1 > 1$, il existe κ_0 et C_2 tels que pour tout $\kappa > \kappa_0$, si

$$pen(m) \geq \frac{\kappa}{n}(\mathcal{D}_m + (1\vee\tau)x_m)$$

alors $\hat{m} = \underset{m\in\tilde{\mathcal{M}}}{\operatorname{argmin}} \left(\gamma_n(\hat{s}_m) + pen(m) + \eta'\right)$ vérifie

$$E(JKL_\rho^{\otimes n}(s^*, \hat{s}_{\hat{m}})) \leq C_1 E\left(\inf_{m\in\tilde{\mathcal{M}}}\left(\inf_{t\in S_m} KL^{\otimes n}(s^*,t)\right) + 2pen(m)\right)$$
$$+ C_2(1\vee\tau)\frac{\Omega^2}{n} + \frac{\eta'+\eta}{n}.$$

D'un point de vue théorique *Massart, Cohen, Le Pennec*

- $\mathcal{S} = (S_m)_{m\in\mathcal{M}}$ la collection de densités conditionnelles,
- $\tilde{\mathcal{S}} = (S_m)_{m\in\tilde{\mathcal{M}}}$ une sous-collection aléatoire de $\mathcal{S}$,
- $(\hat{s}_m)_{m\in\tilde{\mathcal{M}}}$ la collection des η-maxima de vraisemblance dans $\tilde{\mathcal{S}}$.

Si la collection de modèle vérifie les hypothèses K et B_τ, et pour tous les modèles m, les hypothèses H_m et Sep_m,

Alors, pour tout $\rho \in (0,1)$ et $C_1 > 1$, il existe κ_0 et C_2 tels que pour tout $\kappa > \kappa_0$, si

$$pen(m) \geq \frac{\kappa}{n}(\mathcal{D}_m + (1\vee\tau)x_m)$$

alors $\hat{m} = \underset{m\in\tilde{\mathcal{M}}}{\operatorname{argmin}}\left(\gamma_n(\hat{s}_m) + pen(m) + \eta'\right)$ vérifie

$$E(JKL_\rho^{\otimes n}(s^*,\hat{s}_{\hat{m}})) \leq C_1 E\left(\inf_{m\in\tilde{\mathcal{M}}}\left(\inf_{t\in S_m} KL^{\otimes n}(s^*,t)\right) + 2pen(m)\right)$$
$$+ C_2(1\vee\tau)\frac{\Omega^2}{n} + \frac{\eta'+\eta}{n}.$$

D'un point de vue théorique *Massart, Cohen, Le Pennec*

Pour la procédure Lasso-MLE [1]

Si les paramètres sont bornés, il existe $c_1, c_2 > 0$, τ dépendant de s^*, et $C > 0$ tels que, si

$$pen(K,J) \geq c_1 \frac{D_{(K,J)}}{n} \left(1 + c_2 (1 \vee \tau) \log D_{(K,J)}\right),$$

l'estimateur $\hat{s}^{(\hat{K},\hat{J})}$ vérifie

$$E\left(JKL_\rho^{\otimes n}(s^*, \hat{s}^{(\hat{K},\hat{J})})\right) \leq CE\left(\inf_{(K,J)\in\mathcal{K}\times\mathcal{J}} \inf_{t\in\mathcal{S}_{(K,J)}} KL^{\otimes n}(s^*,t) + pen(K,J)\right) + \frac{(1\vee\tau)}{n}$$

[1] Devijver, *Finite mixture regression: a sparse variable selection by model selection for clustering*, preprint, 2014

D'un point de vue théorique *Massart, Cohen, Le Pennec*

Pour la procédure Lasso-Rang [1]

Si les paramètres sont bornés, il existe $c_1, c_2 > 0$, τ dépendant de s^*, et $C > 0$ tels que, si

$$pen(K,J,R) \geq c_1 \frac{D_{(K,J,R)}}{n} \left(1 + c_2(1 \vee \tau) \log D_{(K,J,R)}\right),$$

l'estimateur $\hat{s}^{(\hat{K},\hat{J},\hat{R})}$ vérifie

$$E\left(JKL_\rho^{\otimes n}(s^*, \hat{s}^{(\hat{K},\hat{J},\hat{R})})\right) \leq CE\left(\inf_{(K,J,R)\in\mathcal{K}\times\mathcal{J}\times\mathcal{R}} \inf_{t\in\mathcal{S}_{(K,J,R)}} KL^{\otimes n}(s^*,t) + pen(K,J,R)\right) + \frac{(1\vee\tau)}{n}$$

[1] Devijver, *Joint Rank and Variable Selection for Parsimonious Estimation in High-Dimension Finite Mixture Regression Model*, preprint, 2015

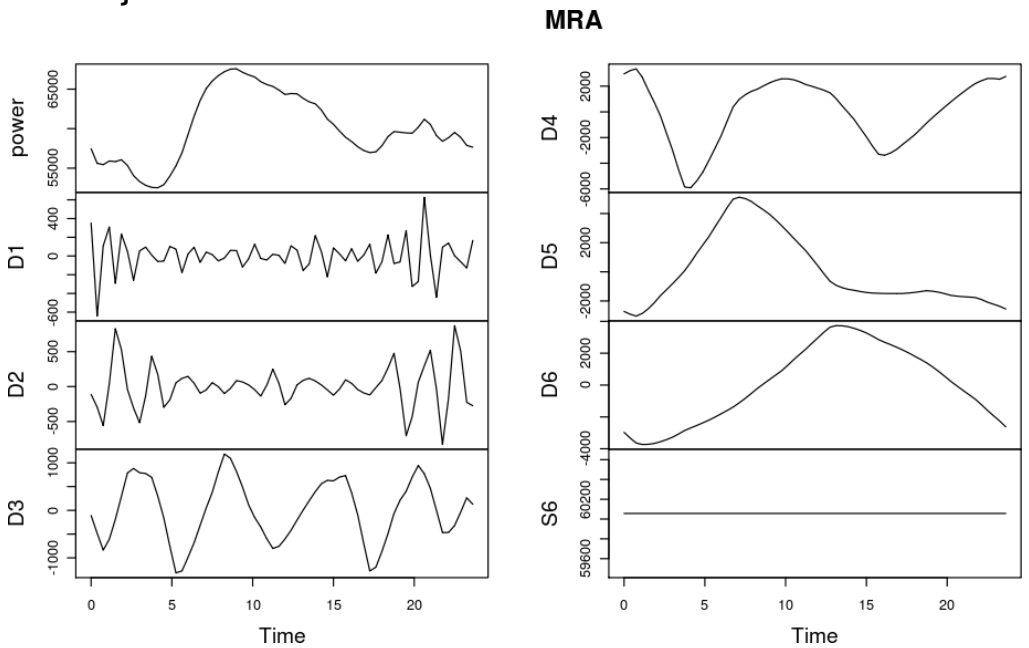
Principe MAP

Pour chaque observation i,

- On estime la probabilité a posteriori $\hat{\tau}_{i,k}$ d'appartenir à chaque classe k ,
- on affecte l'observation à la classe qui majore ces probabilités : $\hat{k}_i = \underset{k \in \{1,\ldots,K\}}{\text{argmax}}\ \hat{\tau}_{i,k}$

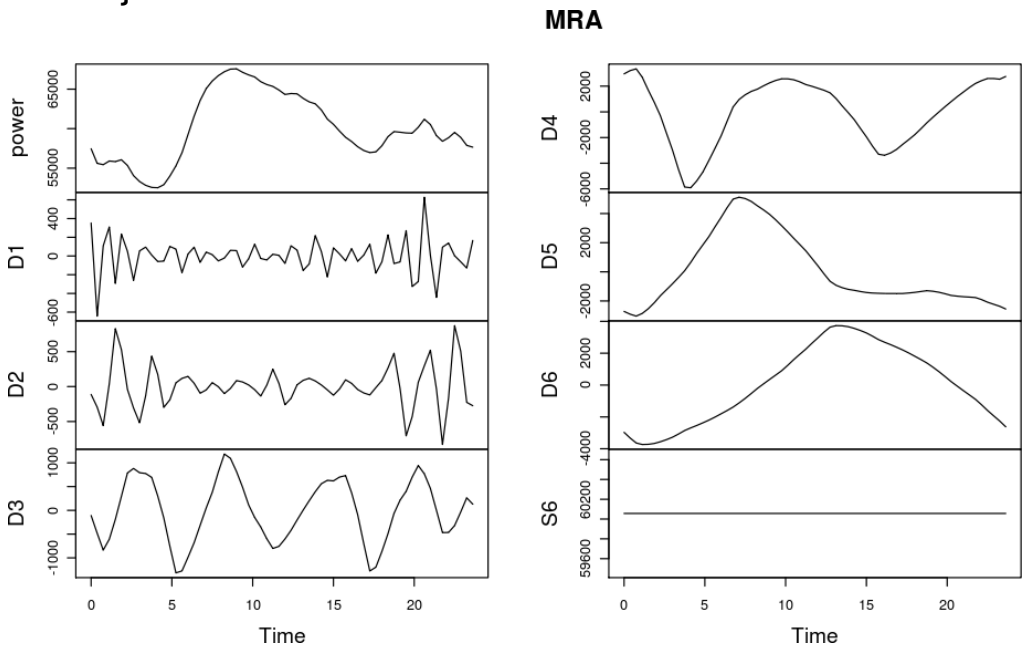
Projection sur une base d'ondelettes

Projection sur une base d'ondelettes



- décomposition hiérarchique
- description d'un signal en terme de tendance générale (partie d'approximation), plus un ensemble de signaux localisés décrits dans les détails.

Projection sur une base d'ondelettes

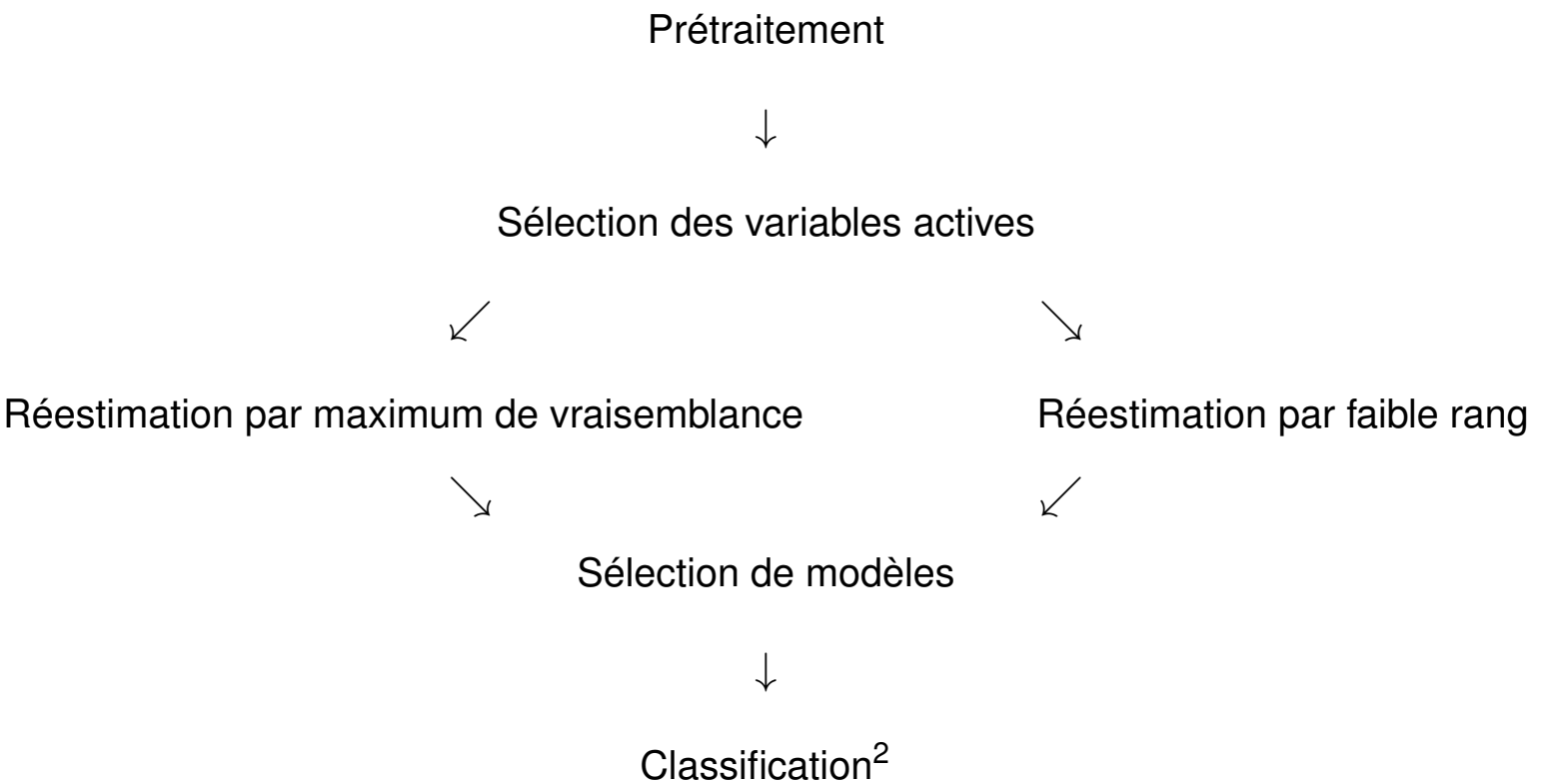


- décomposition hiérarchique
- description d'un signal en terme de tendance générale (partie d'approximation), plus un ensemble de signaux localisés décrits dans les détails.

Si $z \in L_2([0,1])$, on peut écrire

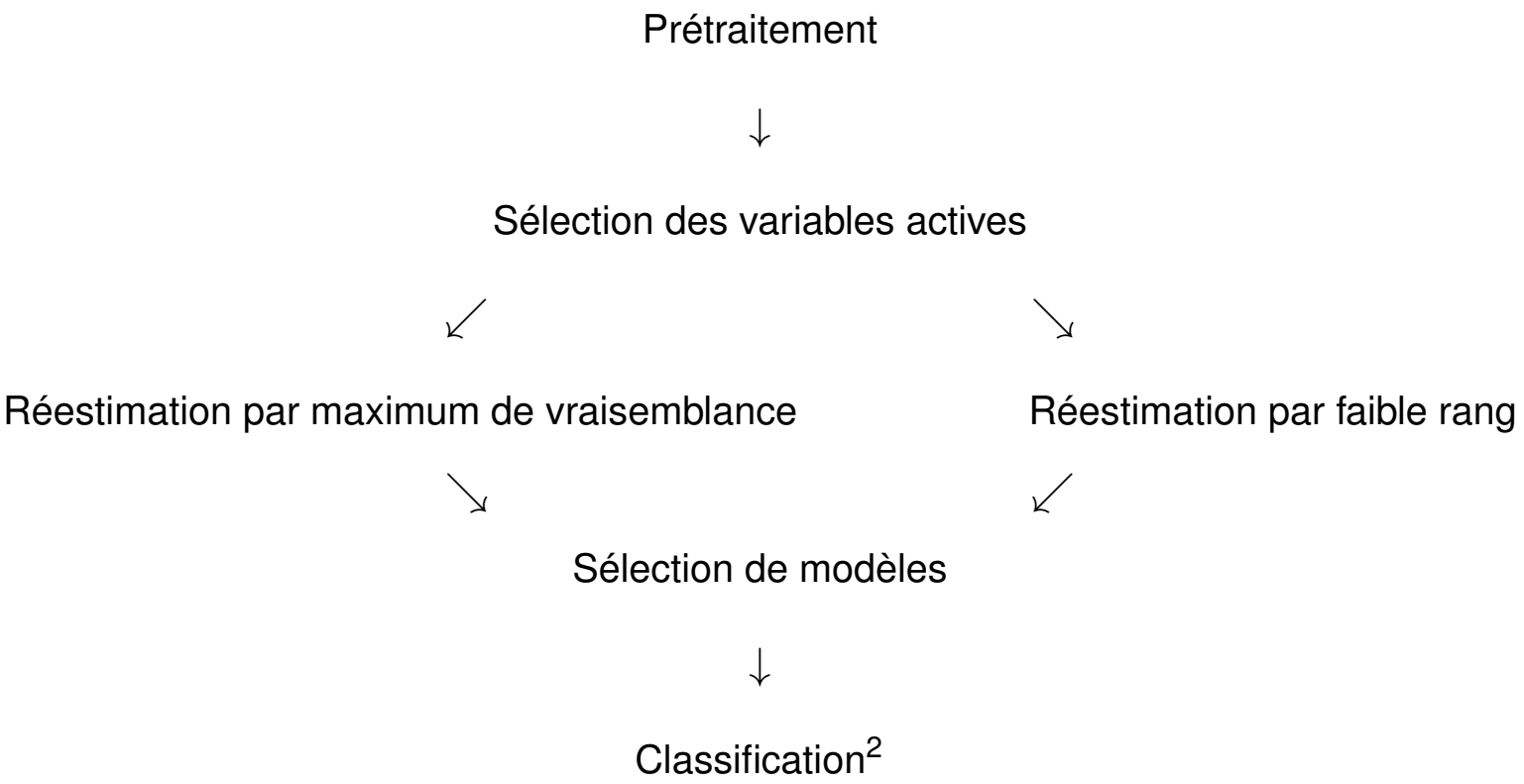
$$z(t) = \sum_{k=0}^{2^{j_0}-1} c_{j_0,k}\phi_{j_0,k}(t) + \sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} d_{j,k}\varphi_{j,k}(t),$$

où $c_{j,k} =< g, \phi_{j,k} >$, $d_{j,k} =< g, \varphi_{j,k} >$ sont respectivement les coefficients d'échelle et les coefficients d'ondelette, et les fonctions ϕ et φ définissent une Analyse Muli-Résolution orthogonale de $L_2([0,1])$.



[2] Devijver, *Model-based clustering for high-dimensional data. Application to functional data*, preprint, 2014

[3] http://www.math.u-psud.fr/~devijver/



Et en pratique ?

Boîte à outils *Selmix* (avec B. Auder) disponible[3] pour Matlab, codée en C, parallélisée.

[2] Devijver, *Model-based clustering for high-dimensional data. Application to functional data*, preprint, 2014

[3] http://www.math.u-psud.fr/~devijver/

Consommation électrique de résidentiels en Irlande [4]

Données

- 4225 consommateurs (résidentiels *ou petites entreprises*)
- consommation observée toutes les 30 min, du 1er Janvier 2010 au 31 Décembre 2010
- accès à des données explicatives (chauffage, vitrage, ...)

Prétraitement / Projection

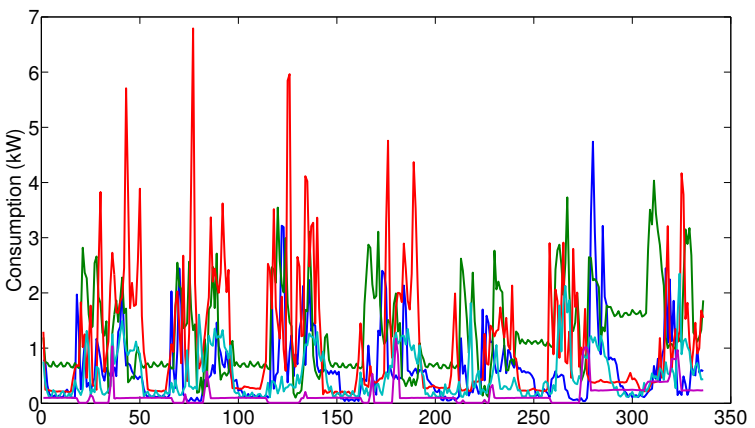


Figure : Échantillon de 5 consommateurs sur une semaine en hiver.

[4] Devijver, Goude et Poggi, *Clustering electricity consumers using high-dimensional regression mixture models*, preprint, 2015

Consommation électrique de résidentiels en Irlande [4]

Données

- 4225 consommateurs (résidentiels *ou petites entreprises*)
- consommation observée toutes les 30 min, du 1er Janvier 2010 au 31 Décembre 2010
- accès à des données explicatives (chauffage, vitrage, ...)

Prétraitement / Projection

2 approches :

- consommation agrégée
- consommation individuelle

[4] Devijver, Goude et Poggi, *Clustering electricity consumers using high-dimensional regression mixture models*, preprint, 2015

On agrège sur les individus. On observe la consommation sur toute l'année : classification des couples de jours.

On agrège sur les individus. On observe la consommation sur toute l'année : classification des couples de jours.

Étape de sélection de modèles

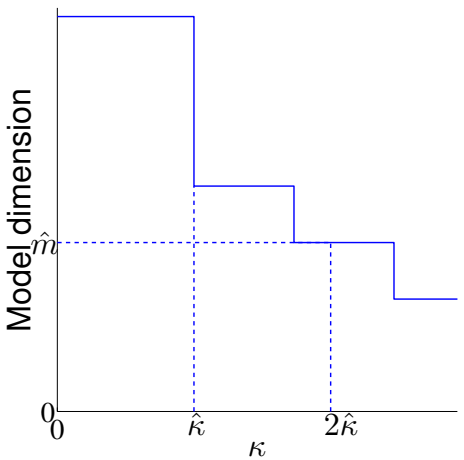


Figure : On choisit le modèle $\hat{m}$ en utilisant l'heuristique des pentes.

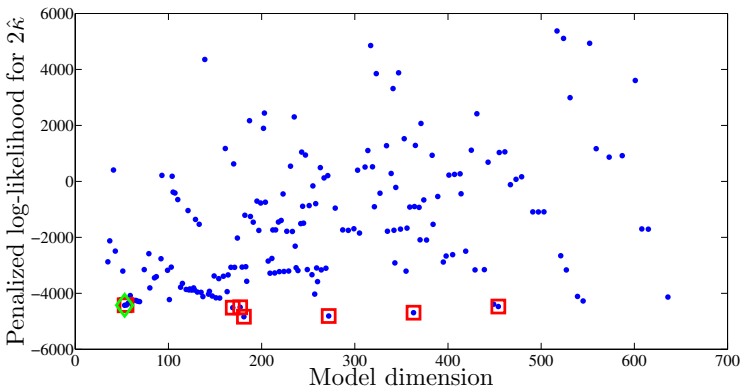


Figure : Minimisation de la logvraisemblance pénalisée. Les modèles intéressants sont encadrés en rouge, le modèle choisi est représenté en vert.

Description des paramètres

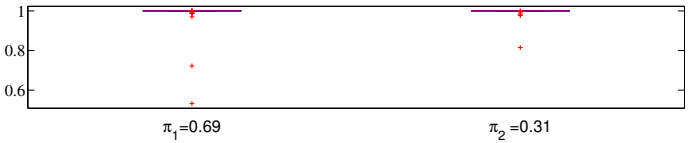


Figure : Proportions dans chaque classe pour le modèle construit par notre procédure

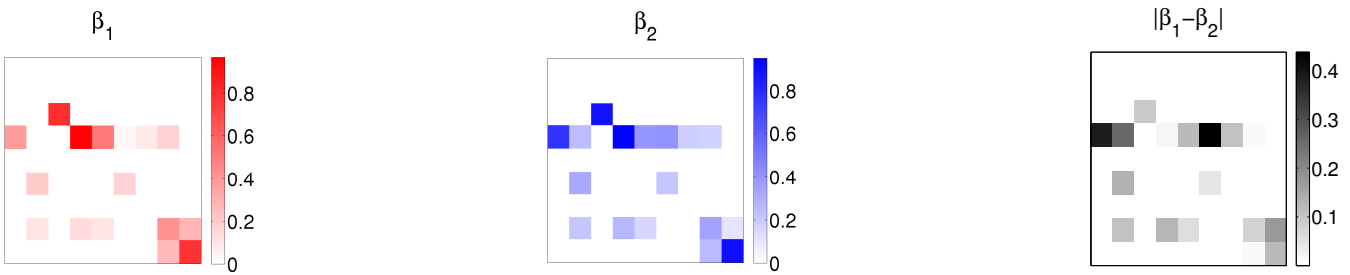


Figure : Pour le modèle choisi, on représente $\hat{\beta}$ dans chaque classe. Les coefficients en valeurs absolues sont représentés dans différents niveaux de couleurs, le blanc correspondant à 0.

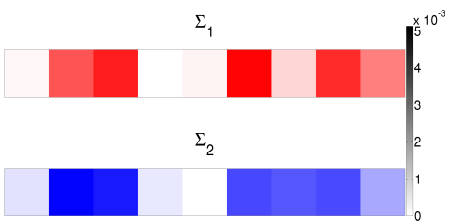


Figure : Pour le modèle choisi, on représente Σ dans chaque classe

Description des groupes

	Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi	Dimanche
	0.88	0.96	0.94	0.98	0.96	0	0
	0.12	0.04	0.06	0.02	0.04	1	1
	0.26	0.46	0.46	0.47	0.51	0	0
	0.1	0.02	0.03	0	0	0.2	0.65
	0.64	0.52	0.5	0.53	0.45	0	0
	0	0	0	0	0.04	0.79	0.35

Table : Pour chaque modèle choisi, on résume la proportion de jour dans chaque classe, et la température moyenne.

Description des groupes

Interprétation	Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi	Dimanche
semaine	0.88	0.96	0.94	0.98	0.96	0	0
week-end	0.12	0.04	0.06	0.02	0.04	1	1
	0.26	0.46	0.46	0.47	0.51	0	0
	0.1	0.02	0.03	0	0	0.2	0.65
	0.64	0.52	0.5	0.53	0.45	0	0
	0	0	0	0	0.04	0.79	0.35

Table : Pour chaque modèle choisi, on résume la proportion de jour dans chaque classe, et la température moyenne.

Description des groupes

Interprétation	Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi	Dimanche
semaine	0.88	0.96	0.94	0.98	0.96	0	0
week-end	0.12	0.04	0.06	0.02	0.04	1	1
semaine, faible T.	0.26	0.46	0.46	0.47	0.51	0	0
week-end, faible T.	0.1	0.02	0.03	0	0	0.2	0.65
semaine, haute T.	0.64	0.52	0.5	0.53	0.45	0	0
week-end, haute T.	0	0	0	0	0.04	0.79	0.35

Table : Pour chaque modèle choisi, on résume la proportion de jour dans chaque classe, et la température moyenne.

Données individuelles

Mardi 5 janvier et mercredi 6 janvier 2010.

On considère un sous-échantillon de 500 consommateurs parmi les 90% les plus proches de la moyenne.

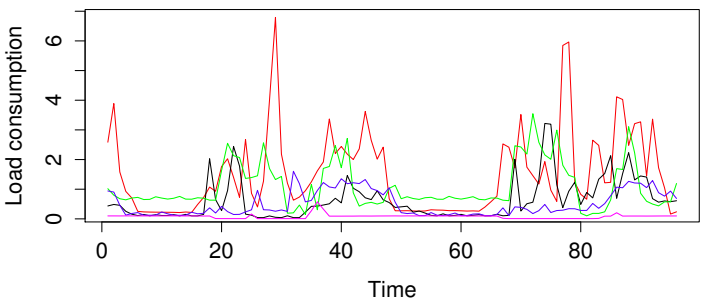


Figure : Échantillon de 5 consommateurs sur une semaine en hiver.

Données individuelles

Mardi 5 janvier et mercredi 6 janvier 2010.

On considère un sous-échantillon de 500 consommateurs parmi les 90% les plus proches de la moyenne.

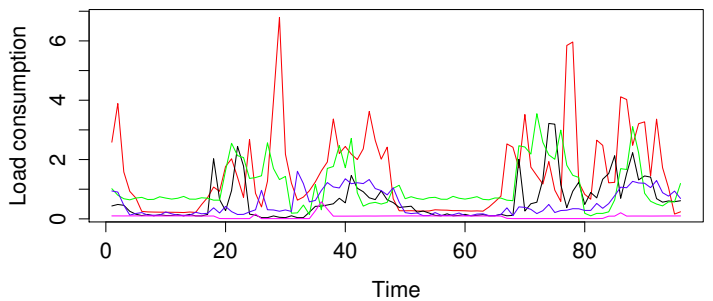


Figure : Échantillon de 5 consommateurs sur une semaine en hiver.

On sélectionne deux modèles.

- M1 : 2 classes
- M2 : 5 classes

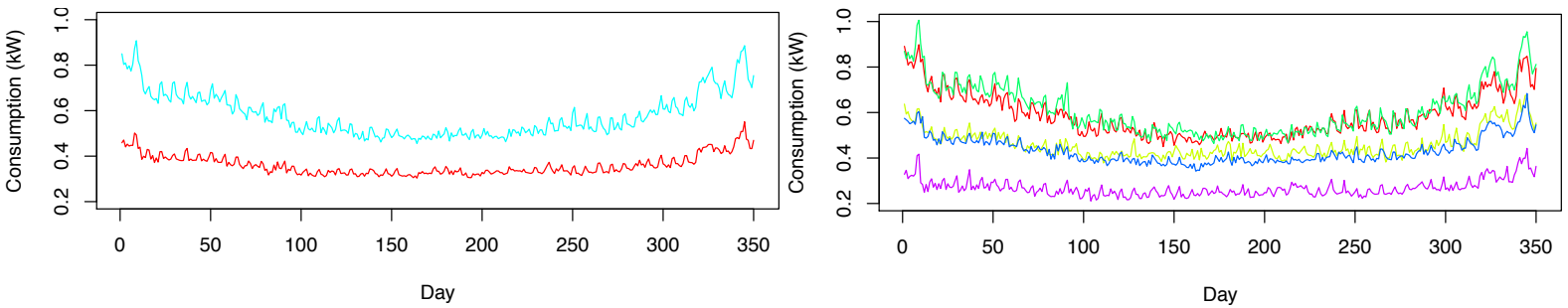


Figure : Consommation moyenne par jour des courbes moyennes par classe sur l'année pour 2 et 5 classes.

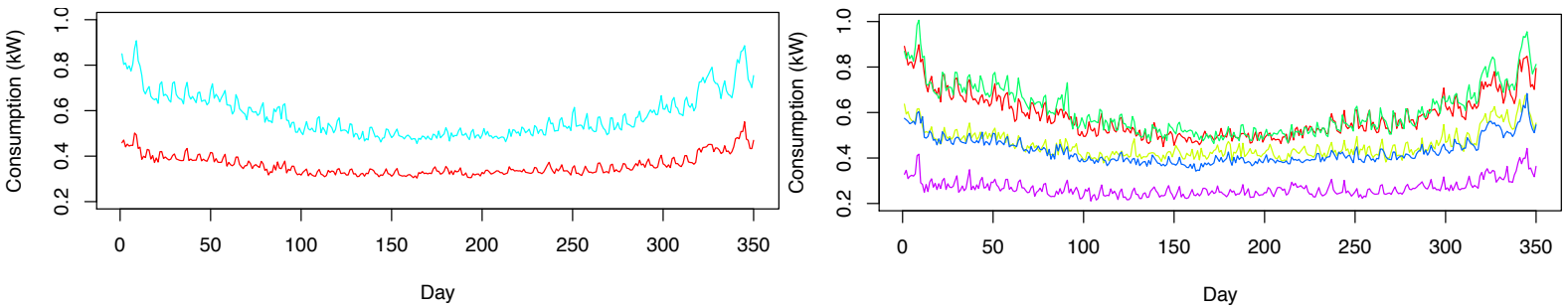


Figure : Consommation moyenne par jour des courbes moyennes par classe sur l'année pour 2 et 5 classes.

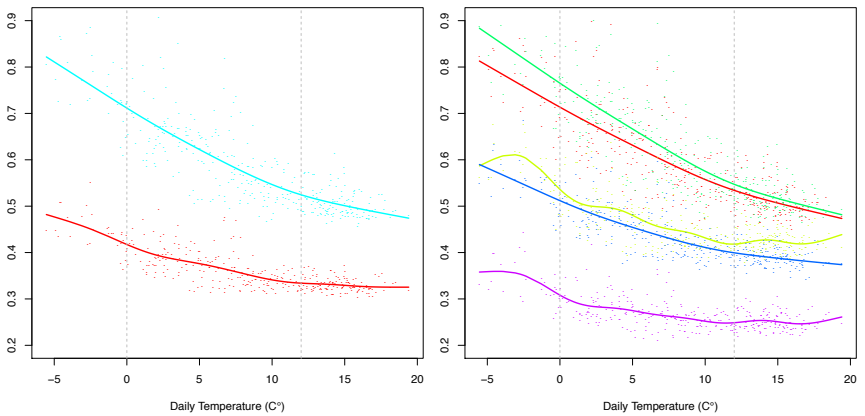


Figure : Consommation moyenne quotidienne des centres des classes en fonction de la température moyenne par jour

Conclusion

- Construction de deux procédures de classification non-supervisée pour des données de régression en grande dimension
- Résultat théorique appuyant l'estimateur du Lasso
- Résultats théoriques appuyant l'étape de sélection de modèles
- Application d'une méthode sur des données réelles

Conclusion

- Construction de deux procédures de classification non-supervisée pour des données de régression en grande dimension
- Résultat théorique appuyant l'estimateur du Lasso
- Résultats théoriques appuyant l'étape de sélection de modèles
- Application d'une méthode sur des données réelles

Perspectives

- D'un point de vue pratique
  - Ajuster un modèle de prédiction dans chaque classe
- D'un point de vue méthodologique
  - Envisager des variables corrélées
  - Améliorer le critère de sélection de modèles dans un but de classification
  - Détection des outliers
- D'un point de vue théorique
  - Borne inférieure
  - Intervalle de confiance
  - Inégalité oracle pour une perte associée à la classification

Bibliographie

- Cohen, S. and Le Pennec, E., *Conditional Density Estimation by Penalized Likelihood Model Selection and Applications*, 2011
- Dempster, A.P. and Laird, N.M. and Rubin, D.B., *Maximum likelihood from incomplete data via the EM algorithm.*, 1977
- Giraud, C., *Low rank multivariate regression*, 2011,
- Massart, P., *Concentration inequalities and model selection*, 2007
- Meynet, C. and Maugis-Rabusseau, C., *A sparse variable selection procedure in model-based clustering*, 2012
- Städler, N. and Bühlmann, P. and van de Geer, S., ℓ_1*-penalization for mixture regression models.*, 2010
- Zhou, N. and Zhu, J., *Group variable selection via a hierarchical lasso and its oracle property*, 2010