

Apprentissage adaptatif en classification et régression

Julien JACQUES

Laboratoire Paul Painlevé - Polytech'Lille
Université Lille 1

22 janvier 2009



Apprentissage statistique¹ :

- observer un phénomène,
- modéliser le phénomène,
- faire des prédictions pour le futur en utilisant le modèle.

⇒ le phénomène n'a **pas évolué** entre l'apprentissage et la prédiction

1. O.Bousquet, S.Boucheron, G.Lugosi (2004)

Apprentissage statistique¹ :

- observer un phénomène,
- modéliser le phénomène,
- faire des prédictions pour le futur en utilisant le modèle.

⇒ le phénomène n'a **pas évolué** entre l'apprentissage et la prédiction

Dans le cas contraire, nous proposons des modèles d'**apprentissage adaptatif** :

- pour la classification, C.Biernacki
- pour la régression, C.Bouveyron

1. O.Bousquet, S.Boucheron, G.Lugosi (2004)

Les modèles d'**apprentissage adaptatif** permettent

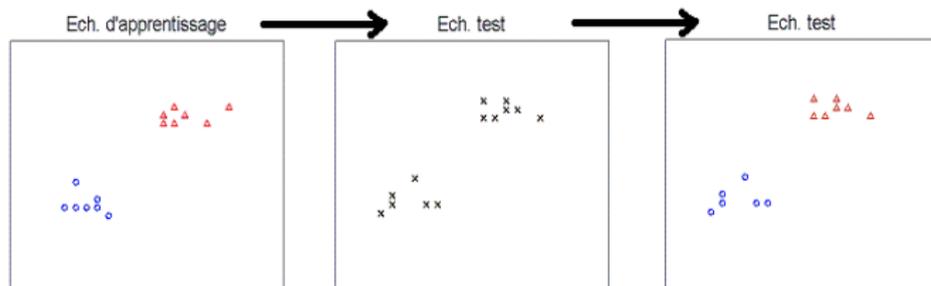
- d'adapter un modèle existant pour une nouvelle situation,
- d'économiser sur le nombre d'observations du nouveau modèle nécessaires.

Exemples d'applications :

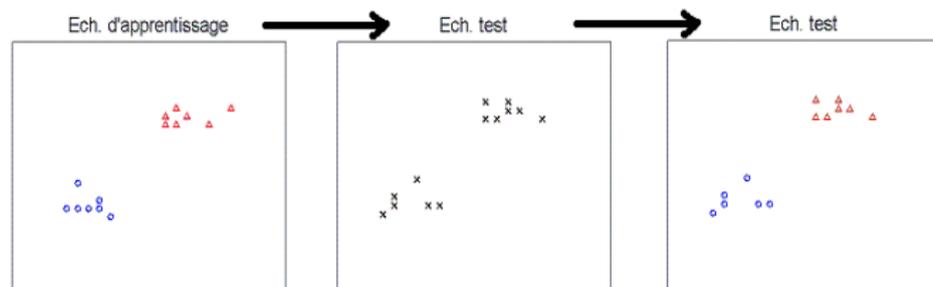
- conquête de nouveaux marchés (Paris \rightarrow province),
- modélisation de phénomènes difficilement observables (sexe de jeunes oiseaux).

- 1 Analyse discriminante généralisée (ADG) gaussienne
- 2 Et lorsque les variables sont binaires ?
 - Définition du modèle d'ADG binaire
 - Estimation des modèles d'ADG binaire
 - Applications en biologie
- 3 Modèles adaptatifs linéaires en régression
 - Modèles adaptatifs en régression
 - Application en biologie : croissance des cellules Tetrahymena
 - Application au marché immobilier des USA

Classification supervisée



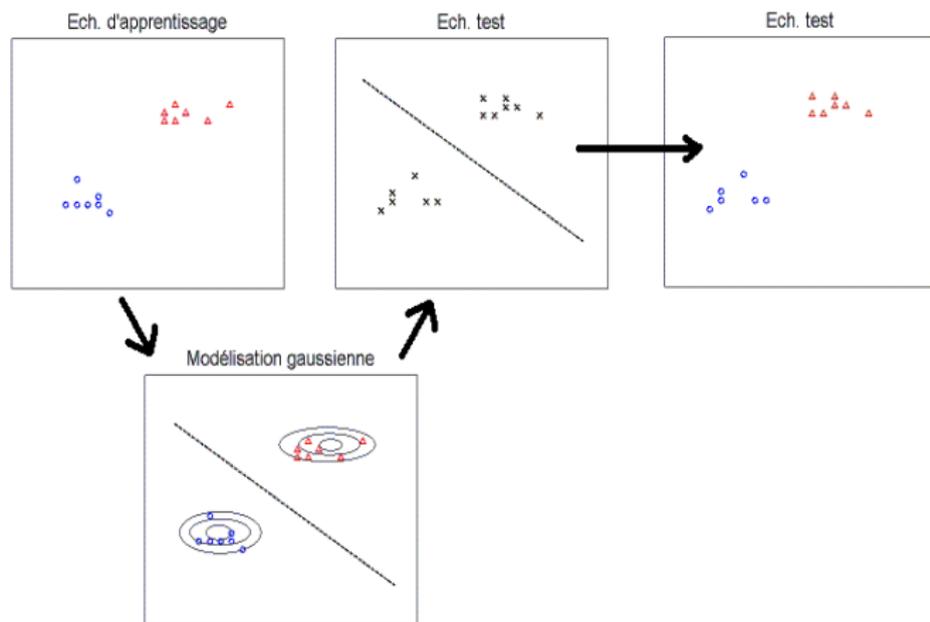
Classification supervisée



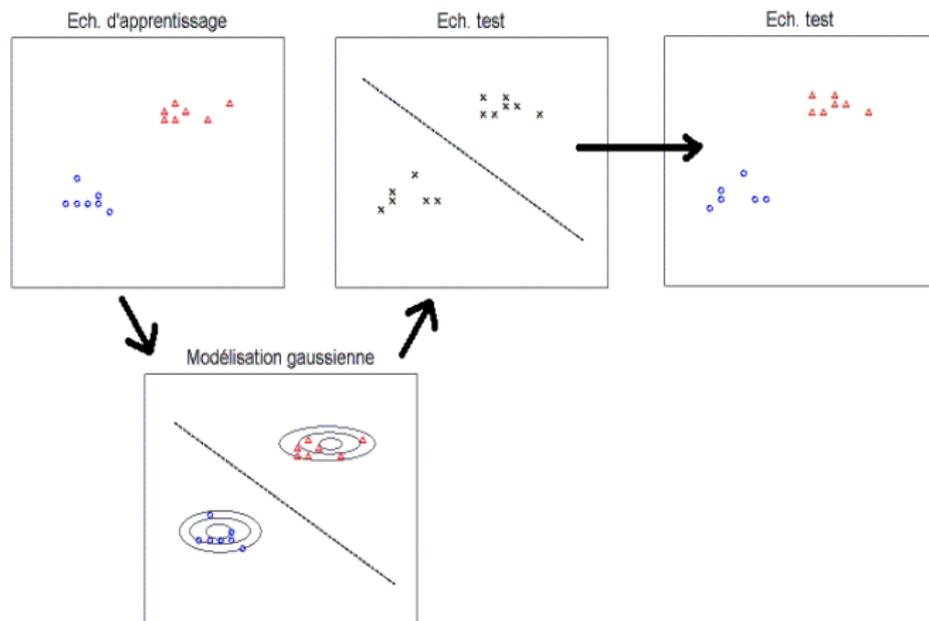
Applications :

- médecine : patients atteints ou non d'une maladie,
- biologie : animaux mâles ou femelles,
- finance (*scoring*) : clients fiables ou à risque...

Analyse discriminante



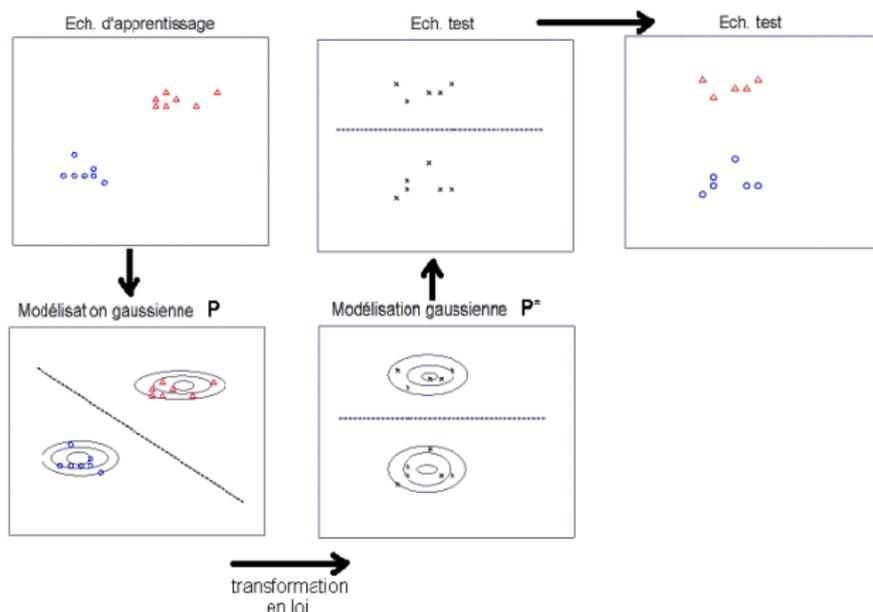
Analyse discriminante



Hypothèse : populations d'apprentissage et de test identiques

Analyse discriminante généralisée (*adaptive*)

Hypothèse : pop. apprentissage $P \neq$ pop. test P^* (*mais liées*).



Analyse discriminante généralisée : applications

Biologie : détermination du sexe d'oiseaux (inconnu sans dissection !)

Espèce *Calanectric diomedea*

P	P^*	lien
<i>borealis</i> (×,atlantique)	<i>diomedea</i> (○,méditerranée)	même espèce

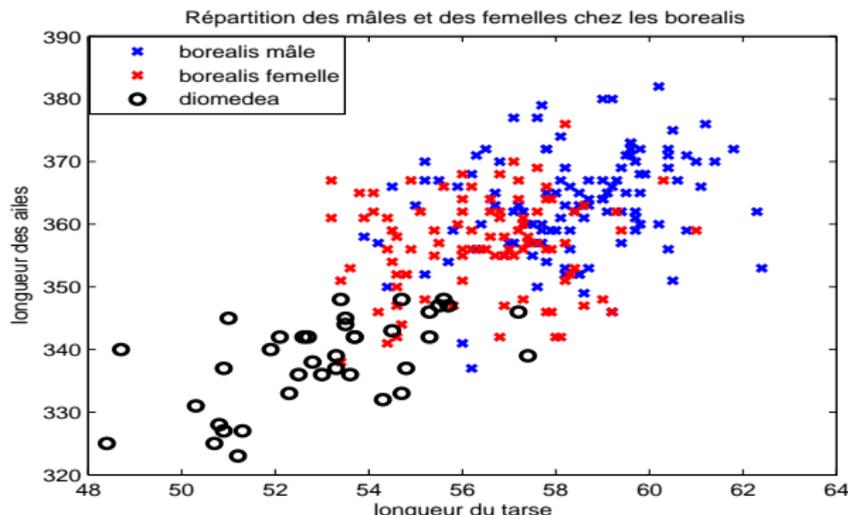


Analyse discriminante généralisée : applications

Biologie : détermination du sexe d'oiseaux (inconnu sans dissection !)

Espèce *Calanectric diomedea*

P	P^*	lien
<i>borealis</i> (×, atlantique)	<i>diomedea</i> (○, méditerranée)	même espèce



Analyse discriminante généralisée : applications

Assurance : détermination des clients à risque

P	P^*	lien
France	Royaume-Uni	même espèce !!

⇒ différence géographique

Analyse discriminante généralisée : applications

Assurance : détermination des clients à risque

P	P^*	lien
France	Royaume-Uni	même espèce !!

⇒ différence géographique

Médecine : détermination des patients à fort risque de second cancer
coll. Registre des Tumeurs CHU Besançon

P	P^*	lien
1970-1990	> 2000	patients d'une même région

⇒ différence temporelle

- 1 Analyse discriminante généralisée (ADG) gaussienne
- 2 Et lorsque les variables sont binaires ?
 - Définition du modèle d'ADG binaire
 - Estimation des modèles d'ADG binaire
 - Applications en biologie
- 3 Modèles adaptatifs linéaires en régression
 - Modèles adaptatifs en régression
 - Application en biologie : croissance des cellules Tetrahymena
 - Application au marché immobilier des USA

Biernacki C., Beninel F. et Bretagnolle V. (2003)

Données :

échantillon S issu de $P : (y_i, z_i)_{i=1\dots n}$, $y_i \in \mathbb{R}^d$, $z_i \in \{0, 1\}^K$

échantillon S^* issu de $P^* : (y_i^*, z_i^*)_{i=1\dots n^*}$, z_i^* **inconnues**.

P	P^*
$Y_{ Z^k=1} \sim \mathcal{N}_d(\mu_k, \Sigma_k)$	$Y_{ Z^{*k}=1}^* \sim \mathcal{N}_d(\mu_k^*, \Sigma_k^*)$
$Z \sim \mathcal{M}(1, p_1, \dots, p_K)$	$Z^* \sim \mathcal{M}(1, p_1^*, \dots, p_K^*)$

Biernacki C., Beninel F. et Bretagnolle V. (2003)

Données :

échantillon S issu de $P : (y_i, z_i)_{i=1\dots n}$, $y_i \in \mathbb{R}^d$, $z_i \in \{0, 1\}^K$

échantillon S^* issu de $P^* : (y_i^*, z_i^*)_{i=1\dots n^*}$, z_i^* **inconnues**.

P	P^*
$Y_{ Z^k=1} \sim \mathcal{N}_d(\mu_k, \Sigma_k)$	$Y_{ Z^{*k}=1}^* \sim \mathcal{N}_d(\mu_k^*, \Sigma_k^*)$
$Z \sim \mathcal{M}(1, p_1, \dots, p_K)$	$Z^* \sim \mathcal{M}(1, p_1^*, \dots, p_K^*)$

Transformation entre populations :

$$Y_{|Z^{*k}=1}^* \sim \phi_k(Y_{|Z^k=1})$$

où $\phi_k : \mathbb{R}^d \rightarrow \mathbb{R}^d$

$$Y_{|Z^{*k}=1}^* \sim \phi_k(Y_{|Z^k=1})$$

où $\phi_k(Y_{|Z^k=1}) = [\phi_k^1(Y_{|Z^k=1}^1), \dots, \phi_k^d(Y_{|Z^k=1}^d)]$ avec $\phi_k^j : \mathbb{R}^d \rightarrow \mathbb{R}$.

Identification de la transformation entre populations

$$Y_{|Z^{*k}=1}^* \sim \phi_k(Y_{|Z^k=1})$$

où $\phi_k(Y_{|Z^k=1}) = [\phi_k^1(Y_{|Z^k=1}^1), \dots, \phi_k^d(Y_{|Z^k=1}^d)]$ avec $\phi_k^j : \mathbb{R}^d \rightarrow \mathbb{R}$.

Hypothèses sur ϕ_k :

- $Y_{|Z^{*k}=1}^{j*}$ ne dépend que de $Y_{|Z^k=1}^j \Rightarrow \phi_k^j : \mathbb{R} \rightarrow \mathbb{R}$

Identification de la transformation entre populations

$$Y_{|Z^{*k}=1}^* \sim \phi_k(Y_{|Z^k=1})$$

où $\phi_k(Y_{|Z^k=1}) = [\phi_k^1(Y_{|Z^k=1}^1), \dots, \phi_k^d(Y_{|Z^k=1}^d)]$ avec $\phi_k^j : \mathbb{R}^d \rightarrow \mathbb{R}$.

Hypothèses sur ϕ_k :

- $Y_{|Z^{*k}=1}^{j*}$ ne dépend que de $Y_{|Z^k=1}^j \Rightarrow \phi_k^j : \mathbb{R} \rightarrow \mathbb{R}$
- ϕ_k^j est $\mathcal{C}^1 \Rightarrow \phi_k^j$ linéaire : $\phi_k^j(t) = d_k^j t + b_k^j$

Identification de la transformation entre populations

$$Y_{|Z^{*k}=1}^* \sim \phi_k(Y_{|Z^k=1})$$

où $\phi_k(Y_{|Z^k=1}) = [\phi_k^1(Y_{|Z^k=1}^1), \dots, \phi_k^d(Y_{|Z^k=1}^d)]$ avec $\phi_k^j : \mathbb{R}^d \rightarrow \mathbb{R}$.

Hypothèses sur ϕ_k :

- $Y_{|Z^{*k}=1}^{j*}$ ne dépend que de $Y_{|Z^k=1}^j \Rightarrow \phi_k^j : \mathbb{R} \rightarrow \mathbb{R}$
- ϕ_k^j est $\mathcal{C}^1 \Rightarrow \phi_k^j$ linéaire : $\phi_k^j(t) = d_k^j t + b_k^j$
- $b_k^j = 0$ (identifiabilité de la transformation)

Identification de la transformation entre populations

$$Y_{|Z^{*k}=1}^* \sim \phi_k(Y_{|Z^k=1})$$

où $\phi_k(Y_{|Z^k=1}) = [\phi_k^1(Y_{|Z^k=1}^1), \dots, \phi_k^d(Y_{|Z^k=1}^d)]$ avec $\phi_k^j : \mathbb{R}^d \rightarrow \mathbb{R}$.

Hypothèses sur ϕ_k :

- $Y_{|Z^{*k}=1}^{j*}$ ne dépend que de $Y_{|Z^k=1}^j \Rightarrow \phi_k^j : \mathbb{R} \rightarrow \mathbb{R}$
- ϕ_k^j est $\mathcal{C}^1 \Rightarrow \phi_k^j$ linéaire : $\phi_k^j(t) = d_k^j t + b_k^j$
- $b_k^j = 0$ (identifiabilité de la transformation)

Transformation entre variables gaussiennes

$$Y_{|Z^{*k}=1}^* \sim D_k Y_{|Z^k=1}$$

avec D_k une matrice diagonale

Règle de classement sur P^*

Si D_1, \dots, D_K sont connues, la règle de classement sur P^* est estimée par *plug-in* :

$$\begin{aligned}\mu_k^* &= D_k \mu_k \\ \Sigma_k^* &= D_k \Sigma_k D_k\end{aligned}$$

Règle de classement sur P^*

Si D_1, \dots, D_K sont connues, la règle de classement sur P^* est estimée par *plug-in* :

$$\begin{aligned}\mu_k^* &= D_k \mu_k \\ \Sigma_k^* &= D_k \Sigma_k D_k\end{aligned}$$

Nombre de paramètres à estimer :

$$\frac{\text{Clustering sur } P^*}{\text{ADG}} \quad \frac{Kd + Kd^2}{Kd}$$

Fondement de l'ADG

Plutôt que d'estimer la règle de classement sur P^* , on estime les matrices D_1, \dots, D_K définissant la transformation entre P et P^*

- 1 Analyse discriminante généralisée (ADG) gaussienne
- 2 Et lorsque les variables sont binaires ?
 - Définition du modèle d'ADG binaire
 - Estimation des modèles d'ADG binaire
 - Applications en biologie
- 3 Modèles adaptatifs linéaires en régression
 - Modèles adaptatifs en régression
 - Application en biologie : croissance des cellules Tetrahymena
 - Application au marché immobilier des USA

- 1 Analyse discriminante généralisée (ADG) gaussienne
- 2 Et lorsque les variables sont binaires ?
 - Définition du modèle d'ADG binaire
 - Estimation des modèles d'ADG binaire
 - Applications en biologie
- 3 Modèles adaptatifs linéaires en régression
 - Modèles adaptatifs en régression
 - Application en biologie : croissance des cellules Tetrahymena
 - Application au marché immobilier des USA

avec C. Biernacki

Données :

échantillon S issu de $P : (x_i, z_i)_{i=1\dots n}$, $x_i \in \{0, 1\}^d$, $z_i \in \{0, 1\}^K$

échantillon S^* issu de $P^* : (x_i^*, z_i^*)_{i=1\dots n^*}$, z_i^* **inconnues**.

P	P^*
$X_{ Z^k=1}^j \sim \mathcal{B}(\alpha_{kj}) \quad \forall j = 1 \dots d$ $Z \sim \mathcal{M}(1, p_1, \dots, p_K)$	$X_{ Z^{*k}=1}^{*j} \sim \mathcal{B}(\alpha_{kj}^*) \quad \forall j = 1 \dots d$ $Z^* \sim \mathcal{M}(1, p_1^*, \dots, p_K^*)$

hyp. des classes latentes : $X_{|Z^k=1}^j \perp X_{|Z^k=1}^{j'} \quad \forall 1 \leq j \leq j' \leq d$.
indépendance conditionnellement à la classe.

Comment identifier une transformation entre variables binaires ?

Idée

supposer que les variables binaires sont issues d'une discrétisation de variables continues (gaussiennes).

Thurstone L.L. 1927, Everitt B.S. 1988

Ex de l'achat d'un produit : le client note ce produit et achète si la note dépasse un certain seuil.

Nous supposons :

- $X_{|Z^k=1}^j$ issue d'une discrétisation de $Y_{|Z^k=1}^j \sim \mathcal{N}(\mu_{kj}, \sigma^2_{kj})$:

$$X_{|Z^k=1}^j = \begin{cases} 0 & \text{si } \lambda_j Y_{|Z^k=1}^j < \lambda_j s_j \\ 1 & \text{sinon} \end{cases} \quad \text{pour } j = 1, \dots, d.$$

où s_j est le seuil de discrétisation

et λ_j est introduit pour ne pas avoir à choisir d'ordre au sein des X^j

- *indépendance conditionnelle* des variables continues sous-jacentes Y^j

Hypothèse de normalité sous-jacente

Nous supposons :

- $X_{|Z^k=1}^j$ issue d'une discrétisation de $Y_{|Z^k=1}^j \sim \mathcal{N}(\mu_{kj}, \sigma^2_{kj})$:

$$X_{|Z^k=1}^j = \begin{cases} 0 & \text{si } \lambda_j Y_{|Z^k=1}^j < \lambda_j s_j \\ 1 & \text{sinon} \end{cases} \quad \text{pour } j = 1, \dots, d.$$

où s_j est le seuil de discrétisation

et λ_j est introduit pour ne pas avoir à choisir d'ordre au sein des X^j

- *indépendance conditionnelle* des variables continues sous-jacentes Y^j

Idée : déduire de la transformation entre les variables continues sous-jacentes $Y_{|Z^k=1}^j$ et $Y_{|Z^k=1}^{*j}$ une transformation entre les variables binaires $X_{|Z^k=1}^j$ et $X_{|Z^k=1}^{*j}$...

Transformation entre populations binaires

... ou encore une transformation entre α_{kj} et α_{kj}^* .

Transformation entre populations binaires

... ou encore une transformation entre α_{kj} et α_{kj}^* .

$$\text{D'après la discrétisation, on a : } \alpha_{kj} = \begin{cases} \Phi\left(\frac{s_j - \mu_{kj}}{\sigma_{kj}}\right) & \text{si } \lambda_j = -1 \\ 1 - \Phi\left(\frac{s_j - \mu_{kj}}{\sigma_{kj}}\right) & \text{si } \lambda_j = 1 \end{cases}$$

avec Φ f.d.r. $\mathcal{N}(0, 1)$.

Transformation entre populations binaires

... ou encore une transformation entre α_{kj} et α_{kj}^* .

D'après la discrétisation, on a : $\alpha_{kj} = \begin{cases} \Phi\left(\frac{s_j - \mu_{kj}}{\sigma_{kj}}\right) & \text{si } \lambda_j = -1 \\ 1 - \Phi\left(\frac{s_j - \mu_{kj}}{\sigma_{kj}}\right) & \text{si } \lambda_j = 1 \end{cases}$
avec Φ f.d.r. $\mathcal{N}(0, 1)$.

$$Y_{|Z^*=k}^* \sim A_k Y_{|Z=k} + b_k \Rightarrow \begin{cases} \mu_{kj}^* = a_{kj} \mu_{kj} + b_{kj} \\ \sigma_{kj}^* = |a_{kj}| \sigma_{kj} \end{cases}$$

identifiabilité $\Rightarrow a_{kj} > 0$,

Transformation entre populations binaires

... ou encore une transformation entre α_{kj} et α_{kj}^* .

D'après la discrétisation, on a : $\alpha_{kj} = \begin{cases} \Phi\left(\frac{S_j - \mu_{kj}}{\sigma_{kj}}\right) & \text{si } \lambda_j = -1 \\ 1 - \Phi\left(\frac{S_j - \mu_{kj}}{\sigma_{kj}}\right) & \text{si } \lambda_j = 1 \end{cases}$
avec Φ f.d.r. $\mathcal{N}(0, 1)$.

$$Y_{|Z^*=k}^* \sim A_k Y_{|Z=k} + b_k \Rightarrow \begin{cases} \mu_{kj}^* = a_{kj} \mu_{kj} + b_{kj} \\ \sigma_{kj}^* = |a_{kj}| \sigma_{kj} \end{cases}$$

identifiabilité $\Rightarrow a_{kj} > 0$,

d'où la transformation entre α_{kj} et α_{kj}^* :

$$\alpha_{kj}^* = \Phi \left(\underbrace{\frac{1}{a_{kj}}}_{\delta_{kj}} \Phi^{-1}(\alpha_{kj}) + \lambda_j \underbrace{\frac{a_{kj} \mu_{kj} + b_{kj} - \mu_{kj}}{a_{kj} \sigma_{kj}}}_{\gamma_{kj}} \right),$$

Transformation entre populations binaires

... ou encore une transformation entre α_{kj} et α_{kj}^* .

D'après la discrétisation, on a : $\alpha_{kj} = \begin{cases} \Phi\left(\frac{s_j - \mu_{kj}}{\sigma_{kj}}\right) & \text{si } \lambda_j = -1 \\ 1 - \Phi\left(\frac{s_j - \mu_{kj}}{\sigma_{kj}}\right) & \text{si } \lambda_j = 1 \end{cases}$
avec Φ f.d.r. $\mathcal{N}(0, 1)$.

$$Y_{|Z^*=k}^* \sim A_k Y_{|Z=k} + b_k \Rightarrow \begin{cases} \mu_{kj}^* = a_{kj} \mu_{kj} + b_{kj} \\ \sigma_{kj}^* = |a_{kj}| \sigma_{kj} \end{cases}$$

identifiabilité $\Rightarrow a_{kj} > 0$,

d'où la transformation entre α_{kj} et α_{kj}^* :

$$\alpha_{kj}^* = \Phi\left(\underbrace{\frac{1}{a_{kj}}}_{\delta_{kj}} \Phi^{-1}(\alpha_{kj}) + \lambda_j \underbrace{\frac{a_{kj} \mu_{kj} + b_{kj} - \mu_{kj}}{a_{kj} \sigma_{kj}}}_{\gamma_{kj}}\right),$$

Estimer α_{kj}^* revient à estimer $\delta_{kj} \in \mathbb{R}^{+*}$, $\lambda_j \in \{-1, 1\}$ et $\gamma_{kj} \in \mathbb{R}$.

Nombre de paramètres (continus) à estimer :

$$\frac{\text{Clustering sur } P^*}{\text{ADG}} \quad \frac{Kd}{2Kd}$$

Règle de classement sur P^*

Nombre de paramètres (continus) à estimer :

Clustering sur P^*	Kd
ADG	$2Kd$

⇒ introduction de modèles de contraintes

			γ_{kj}		
		0	γ	γ_k	γ_j
	1	$[1 \ 0]$	$[1 \ \gamma]$	$[1 \ \gamma_k]$	$[1 \ \gamma_j]$
δ_{kj}	δ	$[\delta \ 0]$	$[\delta \ \gamma]$	$[\delta \ \gamma_k]$	$[\delta \ \gamma_j]$
	δ_k	$[\delta_k \ 0]$	$[\delta_k \ \gamma]$	$[\delta_k \ \gamma_k]$	$[\delta_k \ \gamma_j]$
	δ_j	$[\delta_j \ 0]$	$[\delta_j \ \gamma]$	$[\delta_j \ \gamma_k]$	$[\delta_j \ \gamma_j]$

avec proportions conservées $[p_k \ \delta_j \ 0]$ ou non $[p_k^* \ \delta_j \ 0]$ de P vers P^* .

Choix entre ces modèles : critère BIC

$$BIC = -2l(\hat{\theta}) + \nu \log(N),$$

où

- $l(\hat{\theta})$: maximum de la log-vraisemblance,
- ν : nb de paramètres (continus) du modèle,
- N : taille de l'échantillon.

- 1 Analyse discriminante généralisée (ADG) gaussienne
- 2 Et lorsque les variables sont binaires ?
 - Définition du modèle d'ADG binaire
 - Estimation des modèles d'ADG binaire
 - Applications en biologie
- 3 Modèles adaptatifs linéaires en régression
 - Modèles adaptatifs en régression
 - Application en biologie : croissance des cellules Tetrahymena
 - Application au marché immobilier des USA

L'ADG qui consiste à déduire une règle de classification sur P^* à partir des informations contenues dans S et S^* , nécessite 3 étapes :

- **Étape 1** : estimer p_1, \dots, p_K et α_{kj} du mélange de Bernoulli à partir de l'échantillon S issu de la population P .
- **Étape 2** : estimer p_1^*, \dots, p_K^* ainsi que δ_{kj} , λ_j et γ_{kj} exprimant la relation entre P et P^* , puis en déduire les paramètres α_{kj}^* .
- **Étape 3** : estimer les appartenances aux classes z_1^*, \dots, z_n^* des individus de S^* par *maximum a posteriori*.

L'ADG qui consiste à déduire une règle de classification sur P^* à partir des informations contenues dans S et S^* , nécessite 3 étapes :

- **Étape 1** : estimer p_1, \dots, p_K et α_{kj} du mélange de Bernoulli à partir de l'échantillon S issu de la population P .
- **Étape 2** : estimer p_1^*, \dots, p_K^* ainsi que δ_{kj} , λ_j et γ_{kj} exprimant la relation entre P et P^* , puis en déduire les paramètres α_{kj}^* .
- **Étape 3** : estimer les appartenances aux classes z_1^*, \dots, z_n^* des individus de S^* par *maximum a posteriori*.

Rq : L'ADG est une méthode d'**apprentissage semi-supervisé transductif**

La log-vraisemblance du modèle s'écrit :

$$L(\theta) = \log \left(\prod_{i=1}^{n^*} \sum_{k=1}^K p_k^* \prod_{j=1}^d \alpha_{kj}^{*x_i^{*j}} (1 - \alpha_{kj}^*)^{1-x_i^{*j}} \right),$$

où $\theta = (p_k^*, \delta_{kj}, \lambda_j, \gamma_{kj})$ pour $k = 1, \dots, K$ et $j = 1, \dots, d$.

La log-vraisemblance du modèle s'écrit :

$$L(\theta) = \log \left(\prod_{i=1}^{n^*} \sum_{k=1}^K p_k^* \prod_{j=1}^d \alpha_{kj}^{*x_i^{*j}} (1 - \alpha_{kj}^*)^{1-x_i^{*j}} \right),$$

où $\theta = (p_k^*, \delta_{kj}, \lambda_j, \gamma_{kj})$ pour $k = 1, \dots, K$ et $j = 1, \dots, d$.

- pas de maximum analytique \Rightarrow algorithme itératif,
- données manquantes (appartenances aux classes z_i^* des individus)
 \Rightarrow algorithme EM.

- Initialisation : $\theta^{(0)}$
- Etape E : calcul de $E_{\theta^{(q)}}[L_c(\theta)|x]$
- Etape M : calcul de $\theta^{(q+1)} = \operatorname{argmax}_{\theta} Q(\theta, \theta^{(q)})$

en pratique : plusieurs initialisations aléatoires

Pour info, ici

$$L_c(\theta) = \sum_{i=1}^{n^*} \sum_{k=1}^K z_i^{*k} \log \left(p_k^* \prod_{j=1}^d \alpha_{kj}^* x_i^{*j} (1 - \alpha_{kj}^*)^{(1-x_i^{*j})} \right).$$

- 1 Analyse discriminante généralisée (ADG) gaussienne
- 2 Et lorsque les variables sont binaires ?
 - Définition du modèle d'ADG binaire
 - Estimation des modèles d'ADG binaire
 - Applications en biologie
- 3 Modèles adaptatifs linéaires en régression
 - Modèles adaptatifs en régression
 - Application en biologie : croissance des cellules Tetrahymena
 - Application au marché immobilier des USA

oiseau de mer (*Calanectris diomedea*)

variable discriminante : sexe,

variables descriptives (biométriques) :

longueur bec, tarse, ailes, queue et profondeur bec.

<i>P</i>	<i>P*</i>
<i>borealis</i>	<i>diomedea</i>
atlantique	méditerranée
<i>n</i> = 206	<i>n</i> = 38
45% femelles	58% femelles



pour le besoin de notre étude, ces variables sont discrétisées.

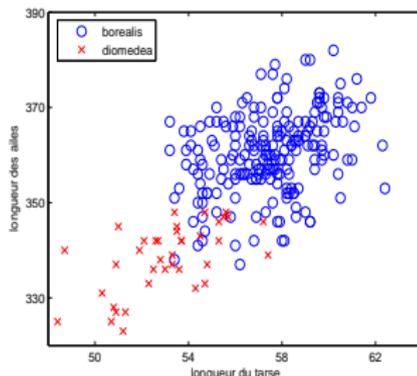
oiseau de mer (*Calanectris diomedea*)

variable discriminante : sexe,

variables descriptives (biométriques) :

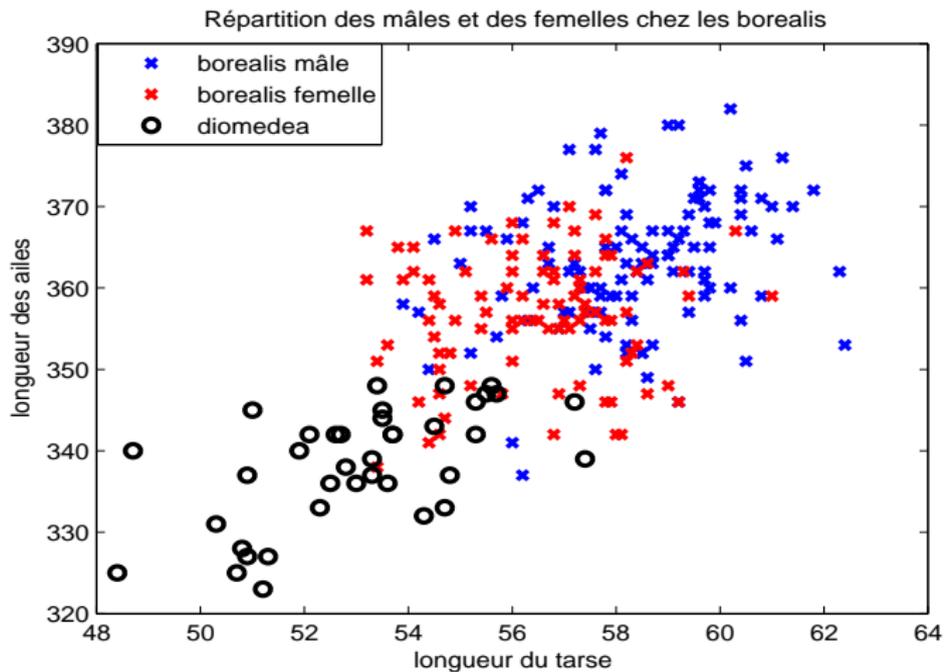
longueur bec, tarse, ailes, queue et profondeur bec.

<i>P</i>	<i>P*</i>
<i>borealis</i>	<i>diomedea</i>
atlantique	méditerranée
<i>n</i> = 206	<i>n</i> = 38
45% femelles	58% femelles

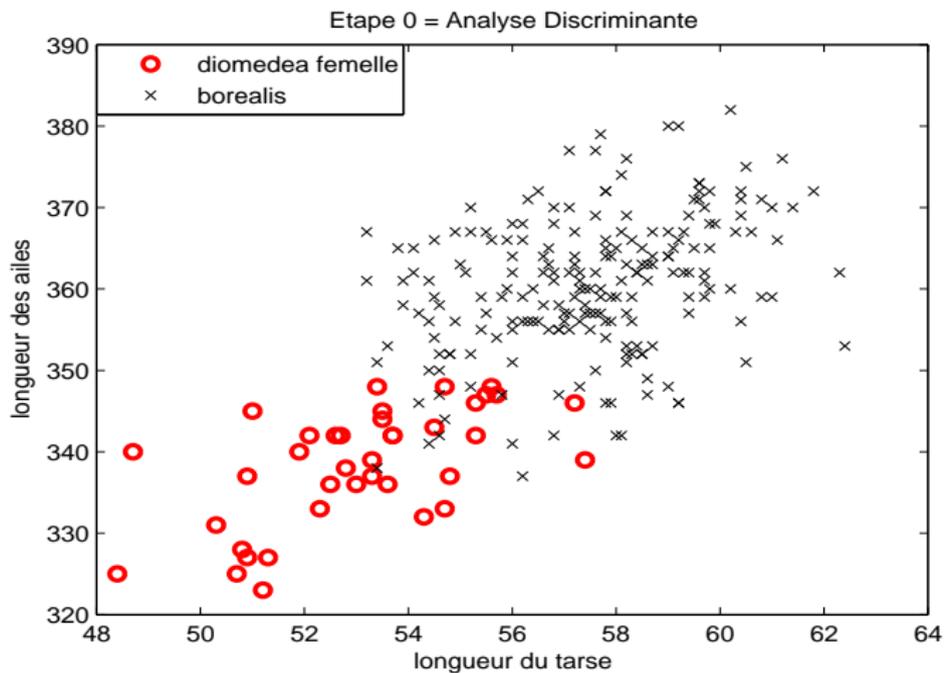


pour le besoin de notre étude, ces variables sont discrétisées.

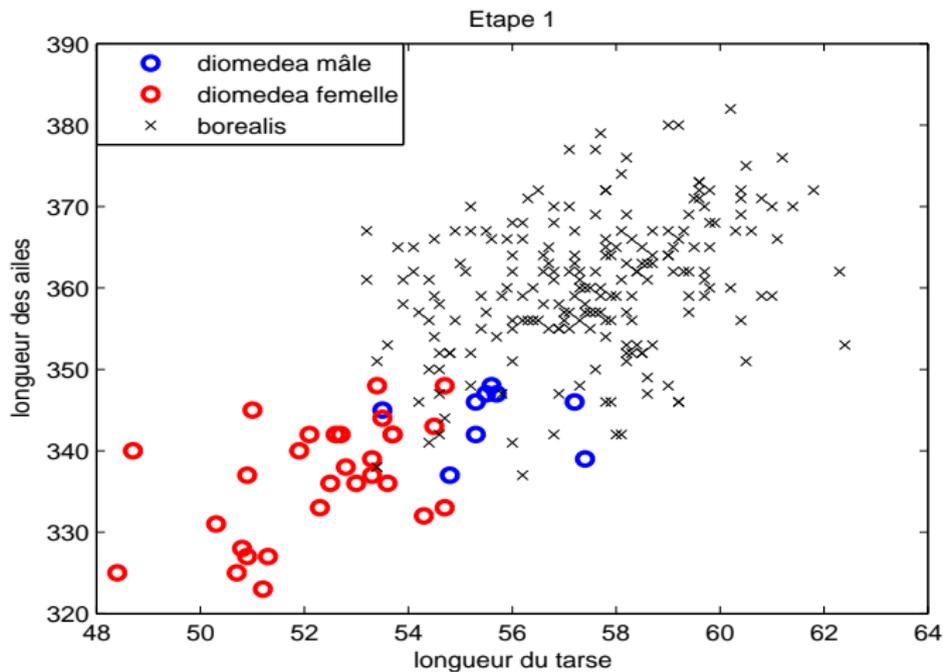
Classement de *Calanectris diomedea*



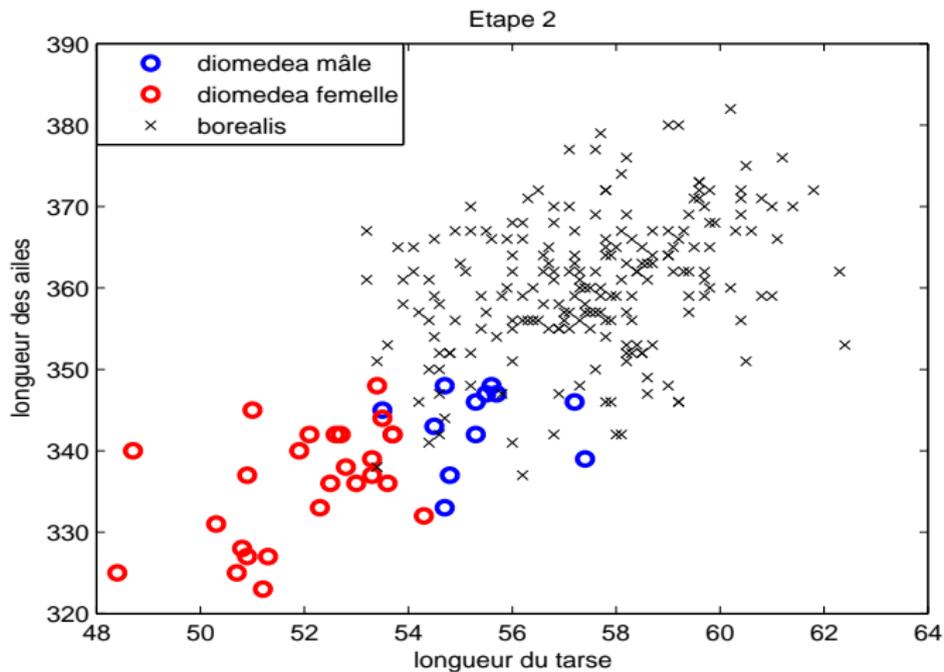
Classement de *Calanectris diomedea*



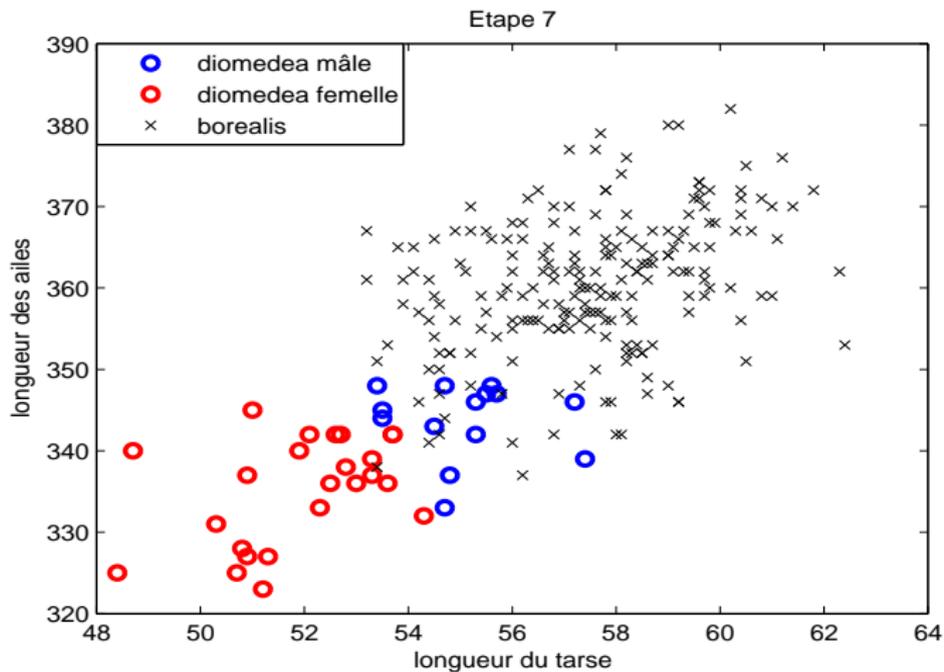
Classement de *Calanectris diomedea*



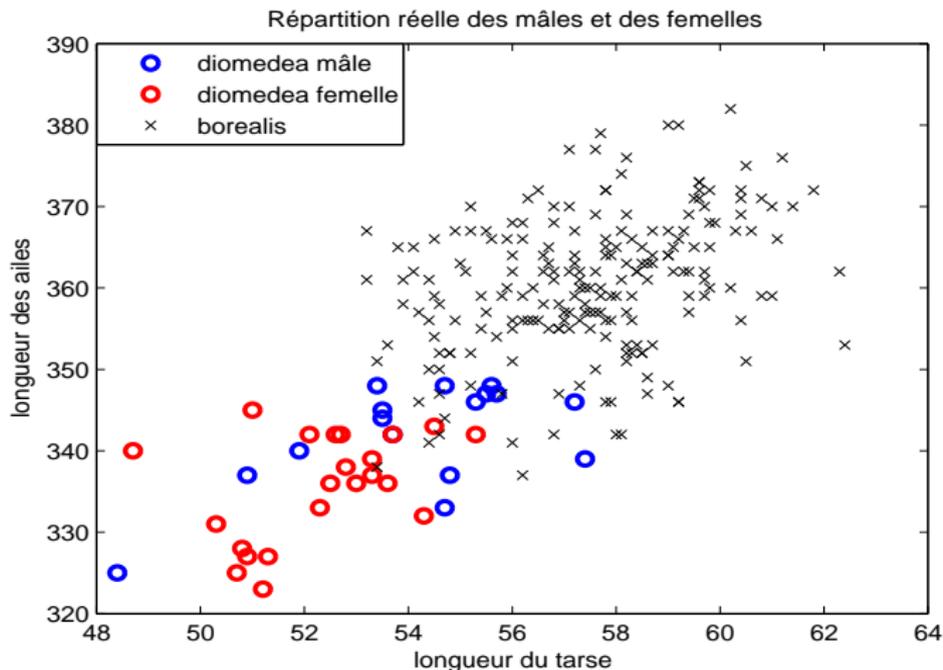
Classement de *Calanectris diomedea*



Classement de *Calanectris diomedea*



Classement de *Calanectris diomedea*



Résultats :

modèle	$[p_k \ 1 \ 0]$	$[p_k \ \delta_j \ \gamma_j]$	$[p_k^* \ \delta \ \gamma]^2$	Classif. autom.	SVM
% erreur	42.1	15.78	21.05	23.7	39.5
nb. erreurs	16	6	8	11	15

2. meilleur modèle suivant BIC

- estimation jointe (intéressante pour de petits échantillons S) :
 - des paramètres de P ,
 - des paramètres de transformation entre P et P^* ,
- classification de données qualitatives (ordinaire ou non),
- GHDDA...

- 1 Analyse discriminante généralisée (ADG) gaussienne
- 2 Et lorsque les variables sont binaires ?
 - Définition du modèle d'ADG binaire
 - Estimation des modèles d'ADG binaire
 - Applications en biologie
- 3 **Modèles adaptatifs linéaires en régression**
 - Modèles adaptatifs en régression
 - Application en biologie : croissance des cellules Tetrahymena
 - Application au marché immobilier des USA

Régression :

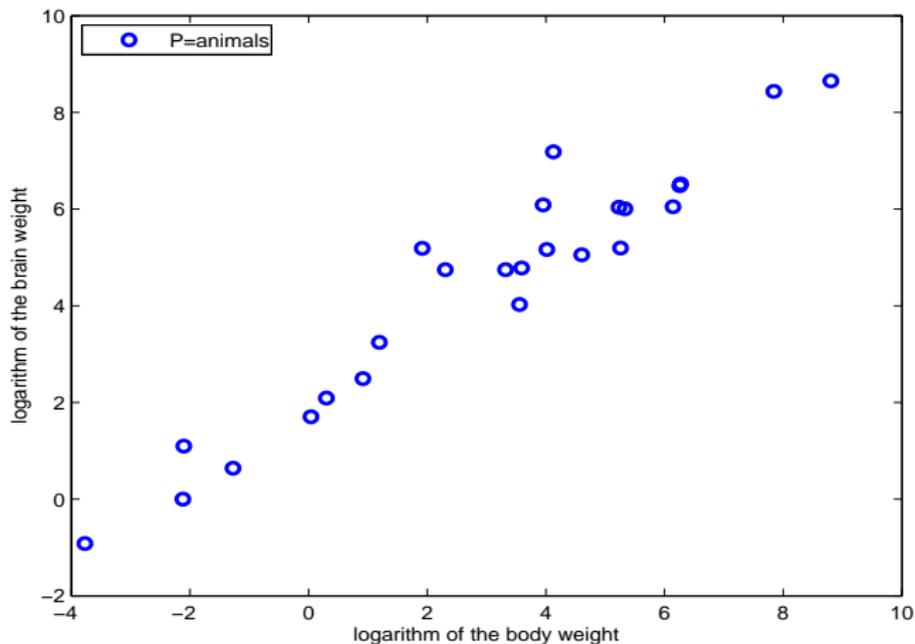
- $Y \in \mathbb{R}$: variable réponse,
- $\mathbf{X} \in \mathbb{R}^p$: variables explicatives,
- identifier une relation entre Y et \mathbf{X} :

$$Y = f(\mathbf{X}, \boldsymbol{\beta}) + \epsilon.$$

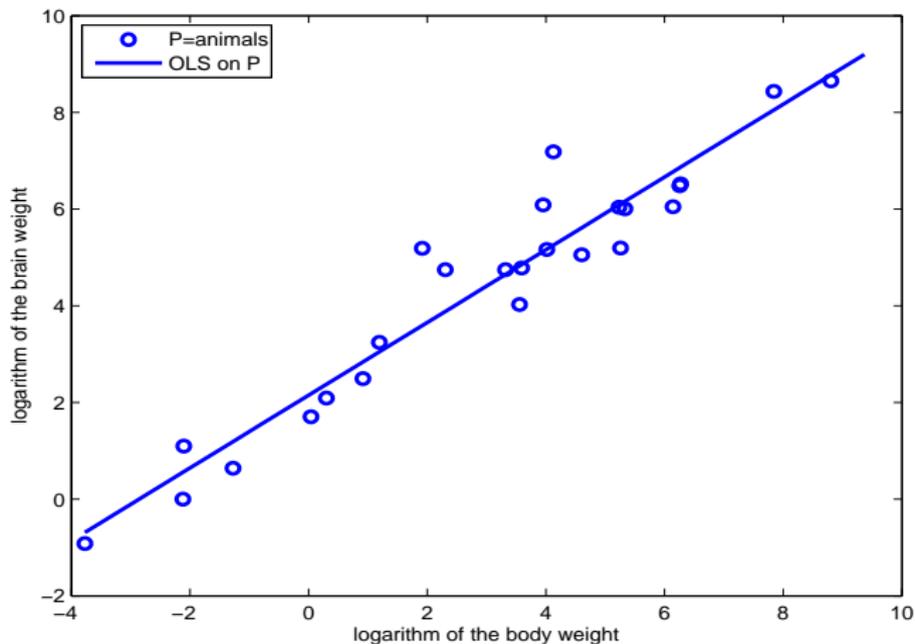
Les principales étapes d'une analyse de régression :

- choisir la fonction de régression : $f(\mathbf{x}, \boldsymbol{\beta}) = \beta_0 + \sum_{i=1}^d \beta_i \psi_i(\mathbf{x})$,
- estimer $\boldsymbol{\beta}$ via l'échantillon d'apprentissage $S = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$,
- prédire y_j^* pour un nouvel échantillon $S^* = \{(\mathbf{x}_1^*, y_1^*), \dots, (\mathbf{x}_{n^*}^*, y_{n^*}^*)\}$.

Un exemple : poids du cerveau vs poids du corps

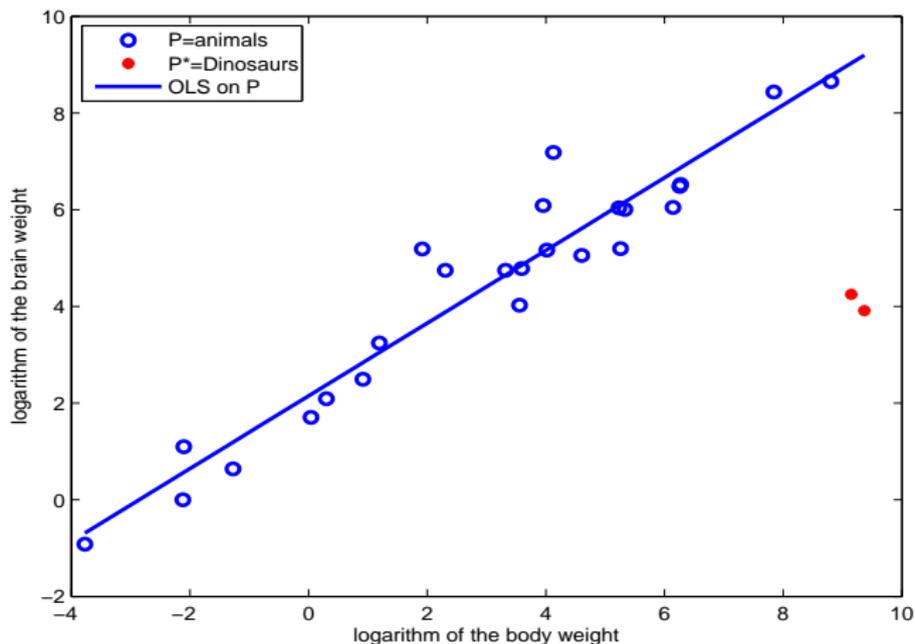


Un exemple : poids du cerveau vs poids du corps



Un exemple : poids du cerveau vs poids du corps

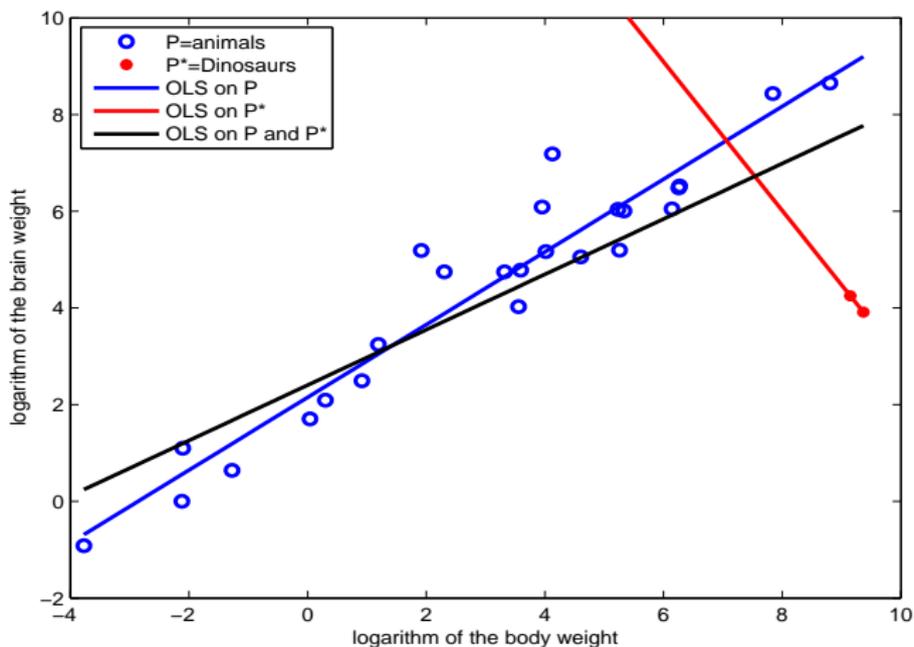
Si S^* est issu d'une population différente de celle de S .



Un exemple : poids du cerveau vs poids du corps

Il faut estimer le nouveau modèle de régression entre Y^* et \mathbf{X}^* :

$$Y^* = f(\mathbf{X}^*, \beta^*) + \epsilon^*.$$



- 1 Analyse discriminante généralisée (ADG) gaussienne
- 2 Et lorsque les variables sont binaires ?
 - Définition du modèle d'ADG binaire
 - Estimation des modèles d'ADG binaire
 - Applications en biologie
- 3 **Modèles adaptatifs linéaires en régression**
 - **Modèles adaptatifs en régression**
 - Application en biologie : croissance des cellules Tetrahymena
 - Application au marché immobilier des USA

Comment adapter un modèle de régression ?

Supposons que le nouvel échantillon S^* est **issu d'une population différente** de celle de S .

Il faut estimer le nouveau modèle de régression entre Y^* et \mathbf{X}^* :

$$Y^* = f(\mathbf{X}^*, \beta^*) + \epsilon^*.$$

hyp. : même base de fonctions de régression

Comment adapter un modèle de régression ?

Supposons que le nouvel échantillon S^* est **issu d'une population différente** de celle de S .

Il faut estimer le nouveau modèle de régression entre Y^* et \mathbf{X}^* :

$$Y^* = f(\mathbf{X}^*, \beta^*) + \epsilon^*.$$

hyp. : même base de fonctions de régression

Plutôt que d'estimer complètement ce nouveau modèle, nous considérons le **lien** suivant **entre les deux fonctions de régression** :

$$\beta^* = \lambda \beta.$$

avec $\lambda = (\lambda_0, \lambda_1, \dots, \lambda_d) \in \mathbb{R}^{d+1}$ le nouveau paramètre.

Complexité (nb. paramètres à estimer) : $(d + 1)$.

Une famille de modèle de transformation

Afin de réduire le nb de paramètres à estimer, nous introduisons des contraintes sur les modèles de transformation

⇒ 7 Modèles Linéaires Adaptatifs :

Modèle	M_0	M_1	M_2	M_3	M_4	M_5	M_6
$\beta_0^* =$	$\lambda_0\beta_0$	β_0	$\lambda_0\beta_0$	$\lambda\beta_0$	β_0	$\lambda_0\beta_0$	β_0
$\beta_i^* =$	$\lambda_i\beta_i$	$\lambda_i\beta_i$	$\lambda\beta_i$	$\lambda\beta_i$	$\lambda\beta_i$	β_i	β_i
Complexité	$d+1$	d	2	1	1	1	0

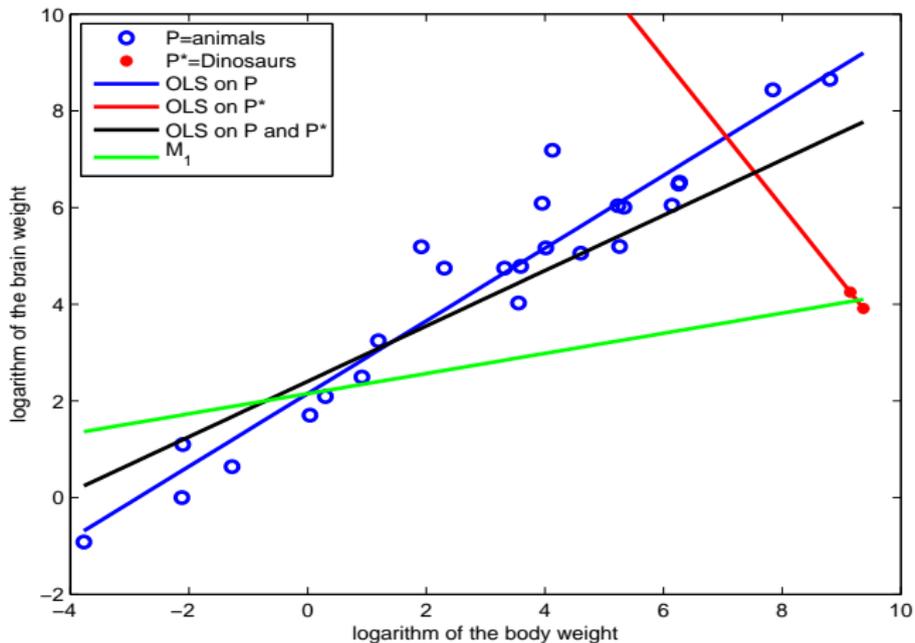
M_0 : revient à apprendre un nouveau modèle de régression sur S^* .

M_6 : suppose que les 2 populations P et P^* sont identiques.

Rq : modèles intermédiaires possibles

Résultat sur l'exemple des dinosaures

Le meilleur modèle est M_1 :



Estimation

La procédure d'estimation est faite de 2 étapes :

- E_1 : estimation de β ,
- E_2 : estimation de β^* *via* l'estimation des paramètres de transformation λ_j .

Ces estimations sont faites par **maximum de vraisemblance** (\sim OLS).

Estimation de M_1 :

$$\mathbf{Y}^* - \beta_0 = \underbrace{\begin{pmatrix} \beta_1 \psi_1(\mathbf{x}_1^*) & \dots & \beta_d \psi_d(\mathbf{x}_1^*) \\ \vdots & \dots & \vdots \\ \beta_1 \psi_1(\mathbf{x}_n^*) & \dots & \beta_d \psi_d(\mathbf{x}_n^*) \end{pmatrix}}_{\mathbf{P}} \underbrace{\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_d \end{pmatrix}}_{\Lambda} + \epsilon^*$$

\Rightarrow l'E.M.V. de Λ est

$$\hat{\Lambda} = (\mathbf{P}'\mathbf{P})^{-1}\mathbf{P}'(\mathbf{Y}^* - \beta_0).$$

En pratique, 2 situations peuvent être rencontrées

- on suppose bien connaître la régression sur P , et on transfère cette information pour estimer une régression sur P^* :
dans ce cas la procédure d'estimation précédente est adéquate.

En pratique, 2 situations peuvent être rencontrées

- on suppose bien connaître la régression sur P , et on transfère cette information pour estimer une régression sur P^* :
dans ce cas la procédure d'estimation précédente est adéquate.
- on veut estimer une régression sur P^* mais n'ayant pas assez de données on utilise celles de P :
⇒ une **estimation jointe** des paramètres de P et P^* est alors plus intéressante.
⇒ algo. itératif :
 - estimation des β
 - estimation des λ

Afin de sélectionner le modèle le plus approprié parmi les 7 Modèles Linéaires Adaptatifs, nous utilisons :

- le critère de validation croisée **PRESS** est utilisé :

$$PRESS = \sum_{j=1}^n \|Y_{(j)}^* - \hat{Y}_{(j)}^*\|_2$$

où

- $\hat{Y}_{(j)}^*$ est la prédiction de Y^* sans utiliser le j -ème individu de S^* ,
 - $Y_{(j)}^*$ est le vecteur Y^* sans le j -ème individu.
- BIC et AIC.

- 1 Analyse discriminante généralisée (ADG) gaussienne
- 2 Et lorsque les variables sont binaires ?
 - Définition du modèle d'ADG binaire
 - Estimation des modèles d'ADG binaire
 - Applications en biologie
- 3 **Modèles adaptatifs linéaires en régression**
 - Modèles adaptatifs en régression
 - **Application en biologie : croissance des cellules Tetrahymena**
 - Application au marché immobilier des USA

Jeu de données *hellung* (package ISwR de R) :

- diamètre moyen des cellules + concentration (nb / ml),
- 2 types de culture : milieu enrichi ou non en glucose.

Jeu de données *hellung* (package ISwR de R) :

- diamètre moyen des cellules + concentration (nb / ml),
- 2 types de culture : milieu enrichi ou non en glucose.

Expérimentation :

- on connaît la **régression diamètre vs concentration** pour le milieu enrichi en glucose,
(*régression polynomiale de degré 3*)

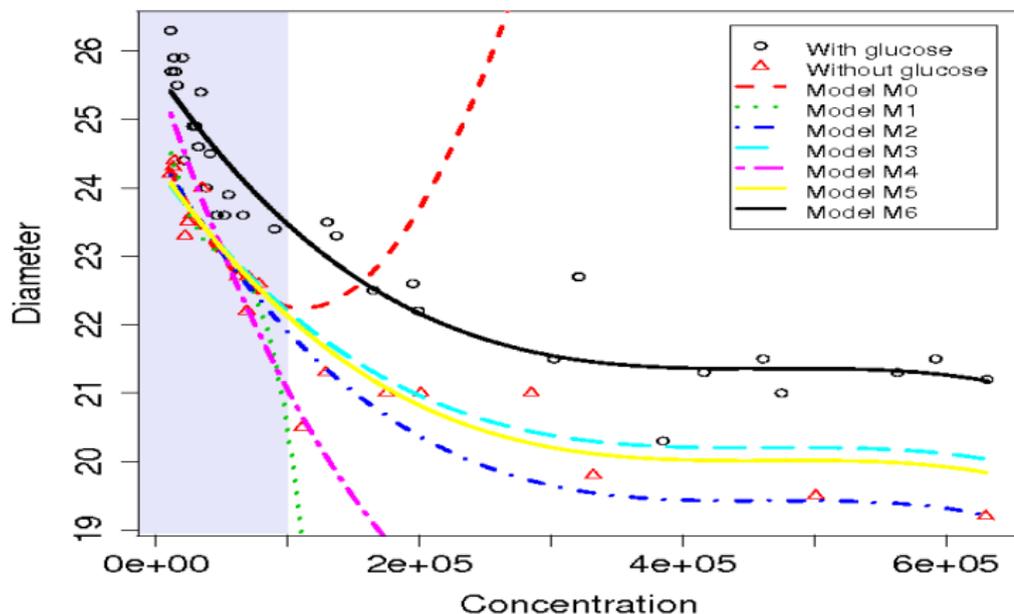
Jeu de données *hellung* (package ISwR de R) :

- diamètre moyen des cellules + concentration (nb / ml),
- 2 types de culture : milieu enrichi ou non en glucose.

Expérimentation :

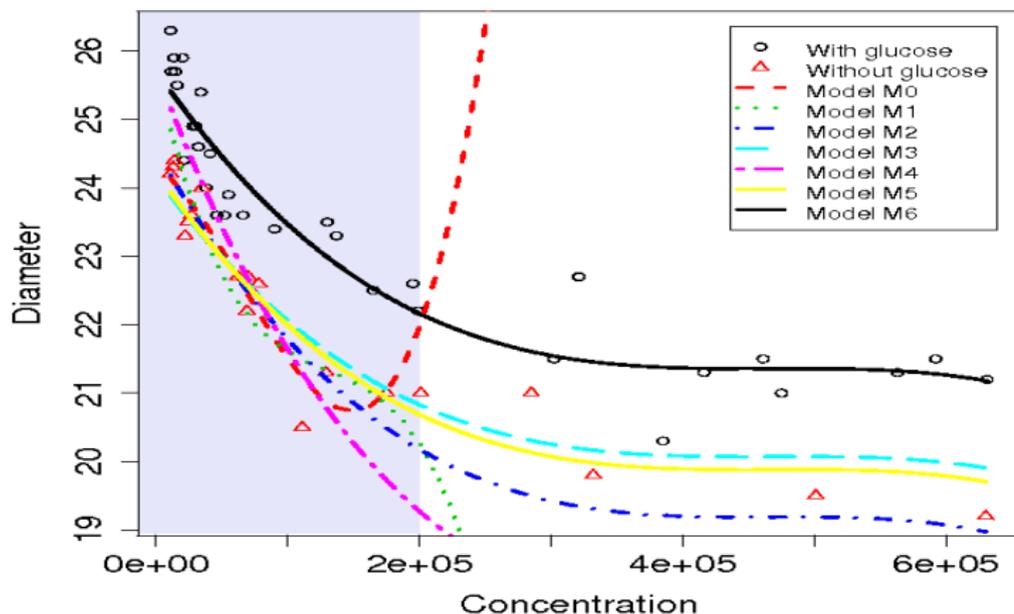
- on connaît la **régression diamètre vs concentration** pour le milieu enrichi en glucose,
(*régression polynomiale de degré 3*)
- on réalise une deuxième culture sans glucose, et on cherche à estimer la régression diamètre vs concentration à **différents instants** de la culture :
 - concentration $\leq 1 \times 10^5$ cel/ml,
 - concentration $\leq 2 \times 10^5$ cel/ml,
 - concentration $\leq 4 \times 10^5$ cel/ml,
 - fin de l'étude.

croissance des cellules Tetrahymena : les résultats



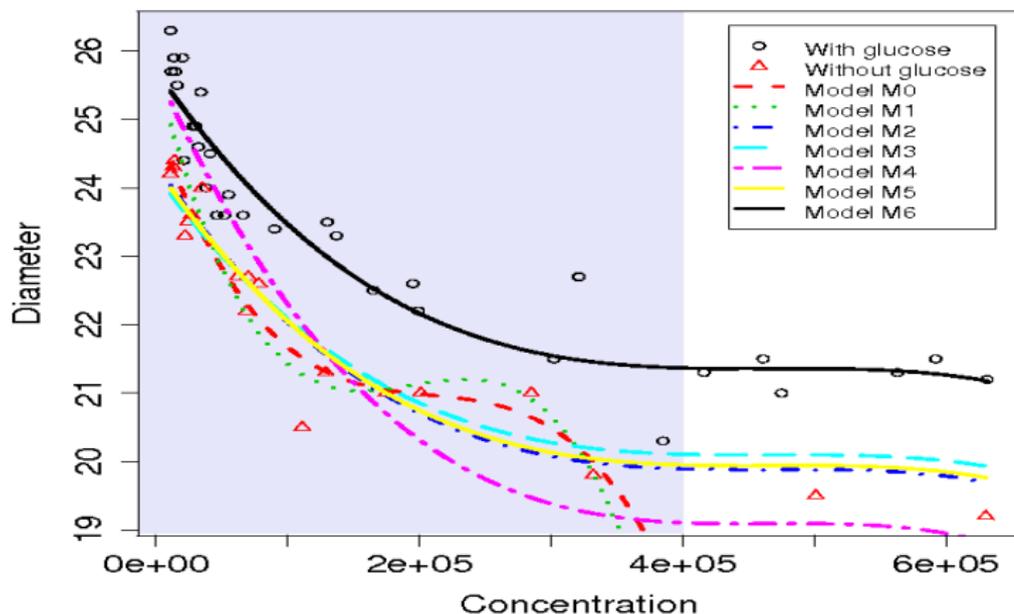
concentration $\leq 1 \times 10^5$ cel/ml

croissance des cellules Tetrahymena : les résultats



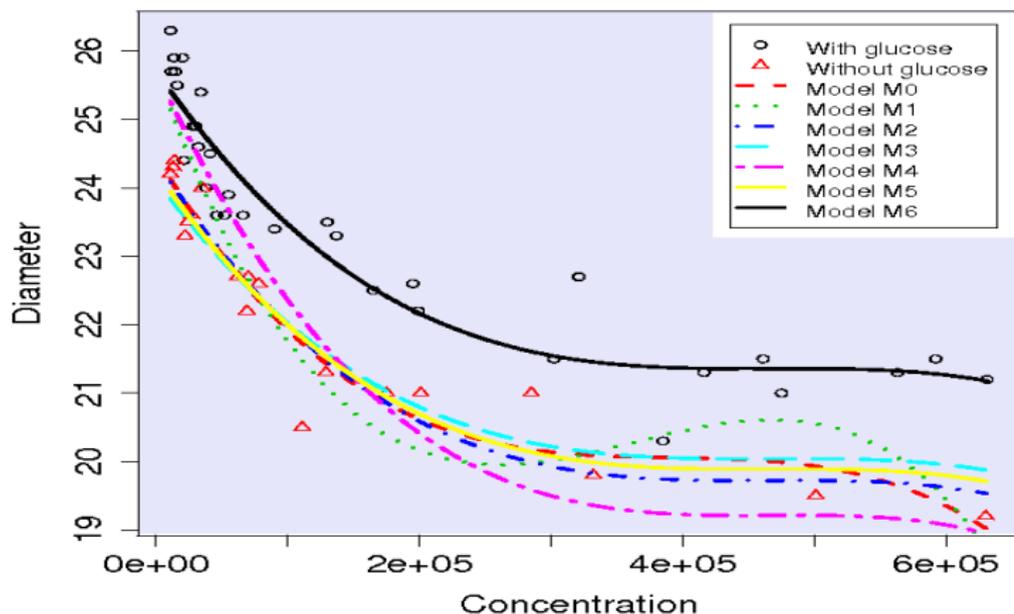
concentration $\leq 2 \times 10^5$ cel/ml

croissance des cellules Tetrahymena : les résultats



concentration $\leq 4 \times 10^5$ cel/ml

croissance des cellules Tetrahymena : les résultats



Echantillon complet

- 1 Analyse discriminante généralisée (ADG) gaussienne
- 2 Et lorsque les variables sont binaires ?
 - Définition du modèle d'ADG binaire
 - Estimation des modèles d'ADG binaire
 - Applications en biologie
- 3 **Modèles adaptatifs linéaires en régression**
 - Modèles adaptatifs en régression
 - Application en biologie : croissance des cellules Tetrahymena
 - **Application au marché immobilier des USA**

Les données proviennent de **1984 American Housing Survey of the U.S. Department of Commerce**.

Nous disposons de :

- la valeur de l'habitation (variable réponse Y),
- 14 variables explicatives X dont
 - le nombre de pièces,
 - la surface,
 - les charges mensuelles,
 - d'autres informations sur l'habitation et les occupants...

marché immobilier des USA : les données

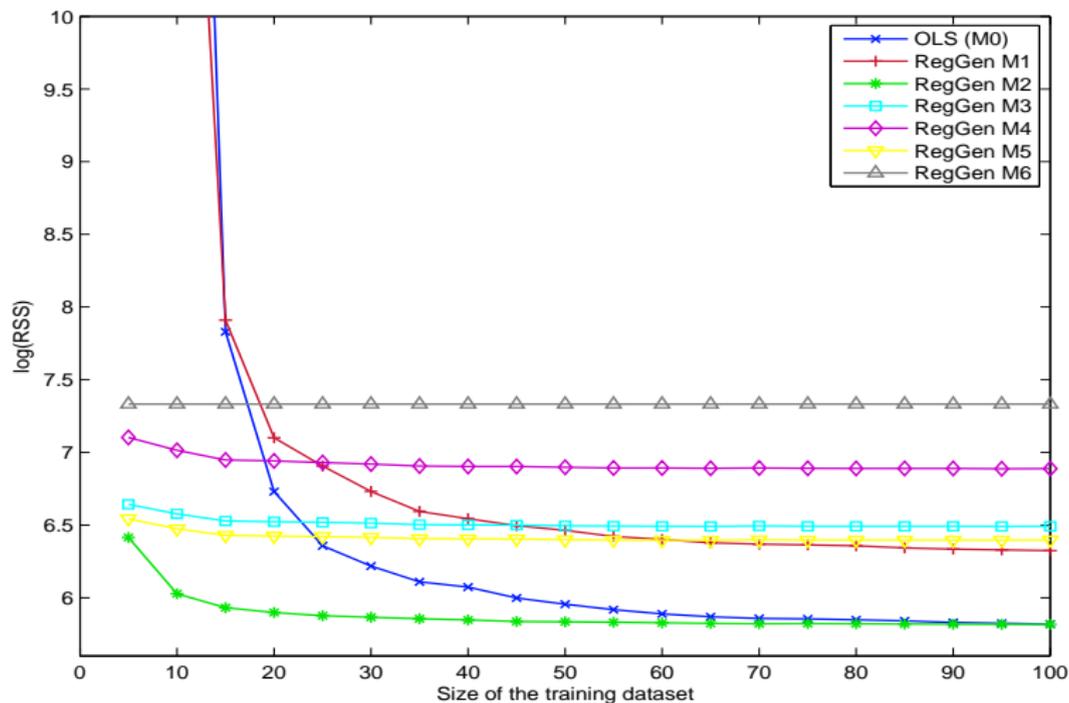
L'objectif de notre application est :

- utiliser une régression préliminaire réalisée sur la ville de **Birmingham** (AL),
- prédire la valeur des habitations de **San José** (CA) en utilisant seulement quelques données.

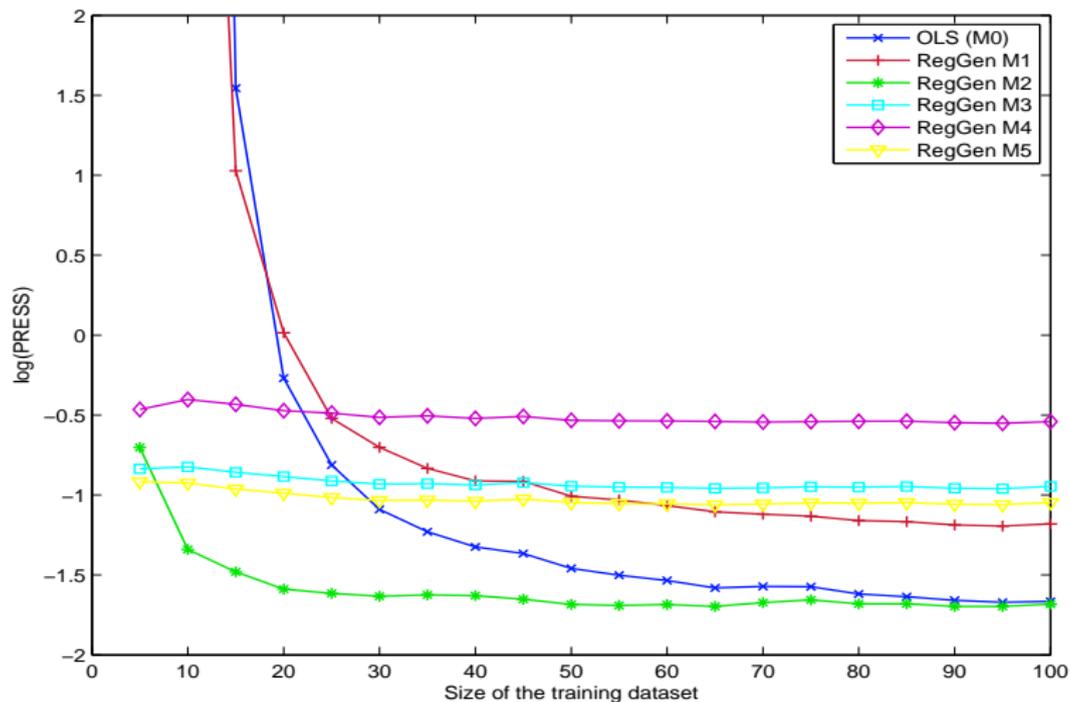


- un **modèle de régression semi-log** est appris sur les 1541 données disponibles pour Birmingham,
- les 7 Modèles Linéaires Adaptatifs sont estimés en utilisant **différentes tailles d'échantillons** pour l'échantillon de données de San José.
choix aléatoire des données, expérience répétée 25 fois.
- pour chaque Modèle Linéaire Adaptatif, la somme des carrés résiduels (**RSS**) et le critère **PRESS** sont calculés.

RSS



PRESS



intérêt pratique

on économise une importante collecte de données,
en recyclant l'information déjà acquise !

- Quelques perspectives :
 - répétabilité de la procédure,
 - modèles mixtes pour données longitudinales,
 - mélanges de régressions.
- preprints + package R dispo sur
<http://math.univ-lille1.fr/~jacques/>