

---

# **Une comparaison de méthodes de discrimination des masses de véhicules automobiles**

A. Rakotomamonjy, R. Le Riche et D. Gualandris

INSA de Rouen / CNRS 1884 et SMS / PSA

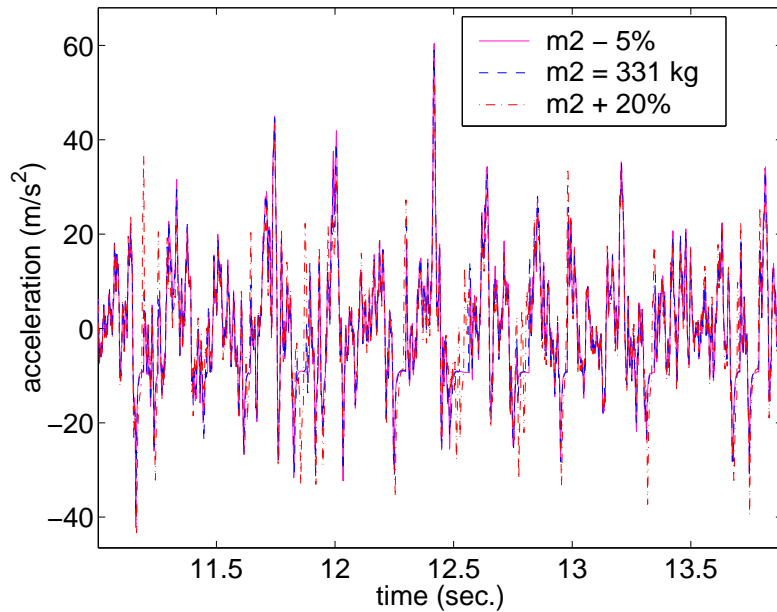
# Enquêtes en clientèle dans l'automobile

---

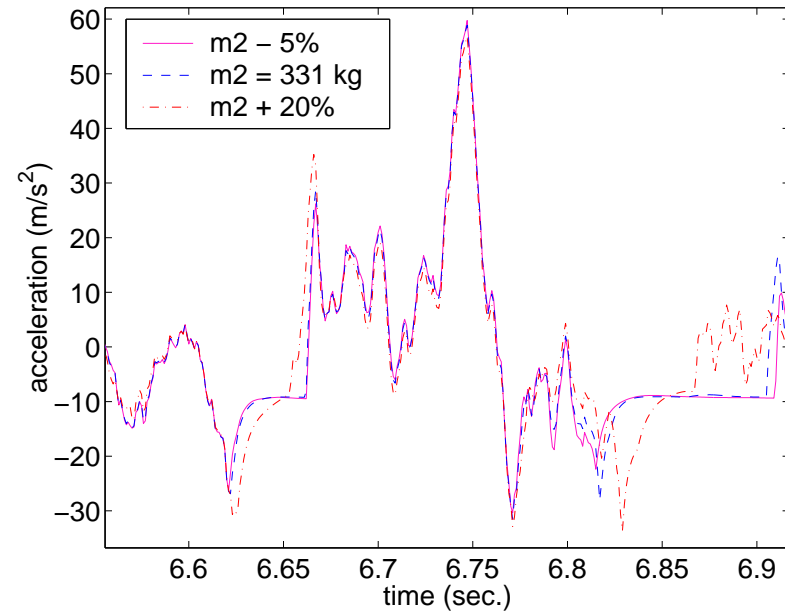
- Dimensionnement fiable des structures automobiles  $\Rightarrow$  enquêtes en clientèle.
- Mesures d'accélération et de vitesse.
- La route est inconnue  $\Rightarrow$  modélisation dynamique impossible.



# Identification de la classe de masse



expl. de signaux



(zoom)

Un problème type : identification de la classe de masse d'un véhicule à partir des signaux mesurés. Puis, identification des efforts transmis, calcul en fatigue des pièces, ....

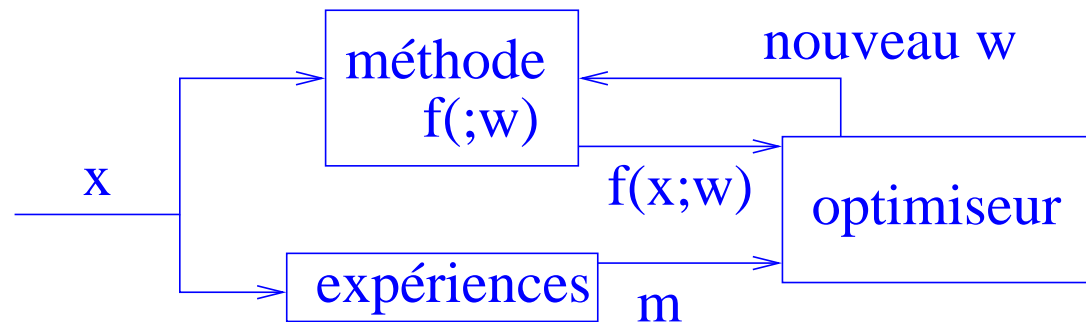
# Données expérimentales

---

- 40 minutes d'enregistrement sur un circuit routier pour trois masses  $m_1$ ,  $m_2$  et  $m_3$ .
- Mesures des accélérations verticales aux 4 pivots de roues + accélérations longitudinales, transverses et verticales au centre de gravité + mesures de vitesses longitudinales.

# Identification de systèmes

---



$$\min_w \text{distance}(m, f(x; w)) ,$$

$x$  et  $m$  connus à travers une base d'expériences de taille finie  $\mathcal{E}$ .

Avantages :

- Seules les entrées et sorties sont nécessaires (utilisable sans modèle physique).
- Offre un cadre probabiliste à la modélisation.
- Une fois identifié, la méthode (classifieur ici) est d'exécution rapide.

# Identification de systèmes (2)

---

Mais :

- Nécessité d'une base expérimentale importante.
- Interprétation physique indirecte : choix des entrées et de la structure du modèle.

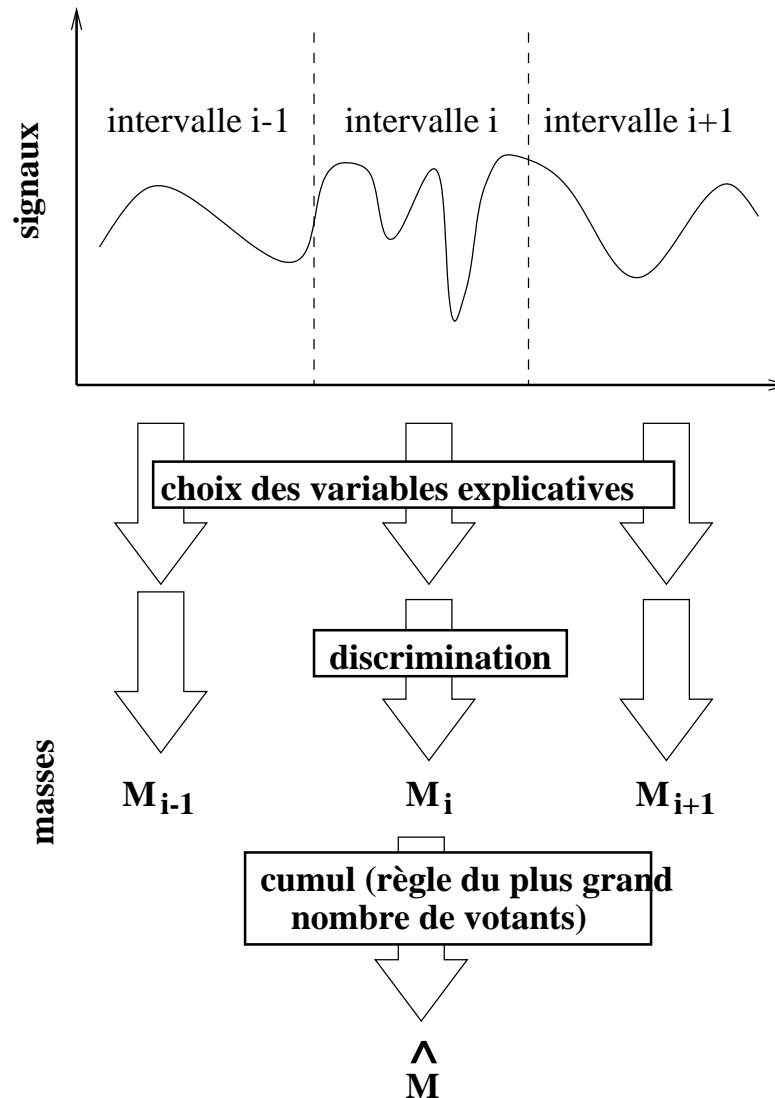
*L. Ljung, "System identification: theory for the user",  
Prentice Hall, 1987.*

*G. Saporta, "Probabilités, analyse des données et statistique",  
Technip, 1990.*

*R. O. Duda, P. E. Hart et D. G. Stork, "Pattern classification",  
John Wiley and Sons, 2000.*

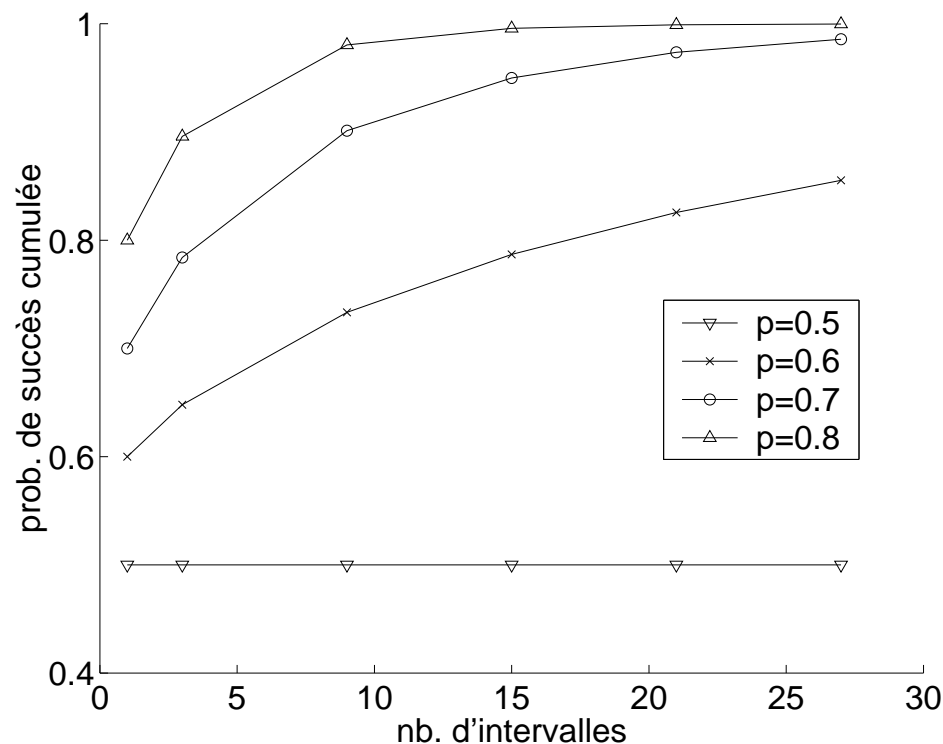
# Principe de la discrimination par masses

- Plus de 100000 points de mesures à traiter pour estimer la masse  $\Rightarrow$  trop de paramètres pour une identification directe.
- Les mesures sont traitées par  $N_i$  intervalles de durée  $D$ .



# Cumul des intervalles

- La masse finale associée aux mesures est celle la plus représentée dans les  $N_i$  intervalles.
- Si la prob. de succès sur 1 intervalle  $p > 0.5$ , la prob. de succès cumulée croît rapidement avec  $N_i$ .





# Variables d'entrées $x$

---

- Les 4 premiers moments statistiques,  $moy$  (vit. uniquement),  $var$ ,  $skew$ ,  $kurt = \frac{1}{D} \sum_{i=1}^D a_i^j \quad j = 1, 4$ .
- La moyenne des minima et maxima des signaux,  $a_{min}$  et  $a_{max}$ , chaque extremum étant relatif à un horizon de  $\pm 1$  sec.
- L'énergie du signal sur la plage de fréquence  $F$ ,  
 $E = \sum_{i|f_i \in F} DSP(f_i)^2$ .
- Les moments relatifs d'ordre 1 à 4 de la DSP,  
 $M_j = \frac{1}{E} \sum_{i|f_i \in F} f_i^j DSP(f_i), j = 1, 4$ .

× 7 signaux d'accélérations et 1 de vitesses,

⇒  $d = 81$  variables d'entrées potentielles.

# Des optimisations à 4 niveaux

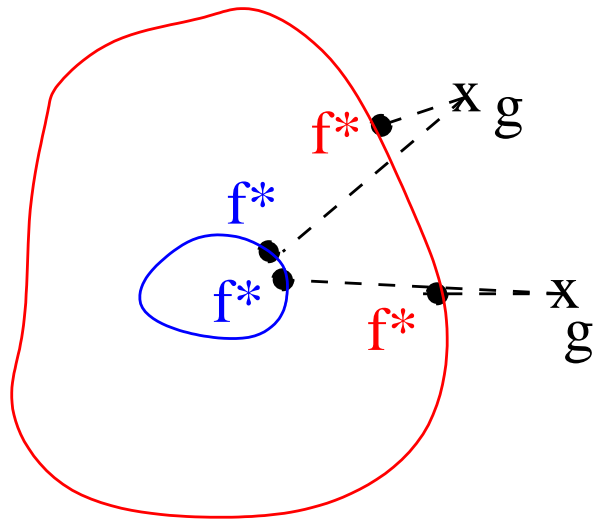
---

On cherche le meilleur jeu [variables d'entrées / classeur] pour identifier la masse. Il y a 4 niveaux d'optimisation :

1. Choix des variables explicatives (définition de l'espace de recherche).
2. Choix d'un critère de discrimination (d'une méthode ou classeur).
3. Choix d'un niveau de complexité du classeur.
4. Choix des paramètres (autres que ceux qui contrôlent la complexité).

# Complexité et variance d'un classeur

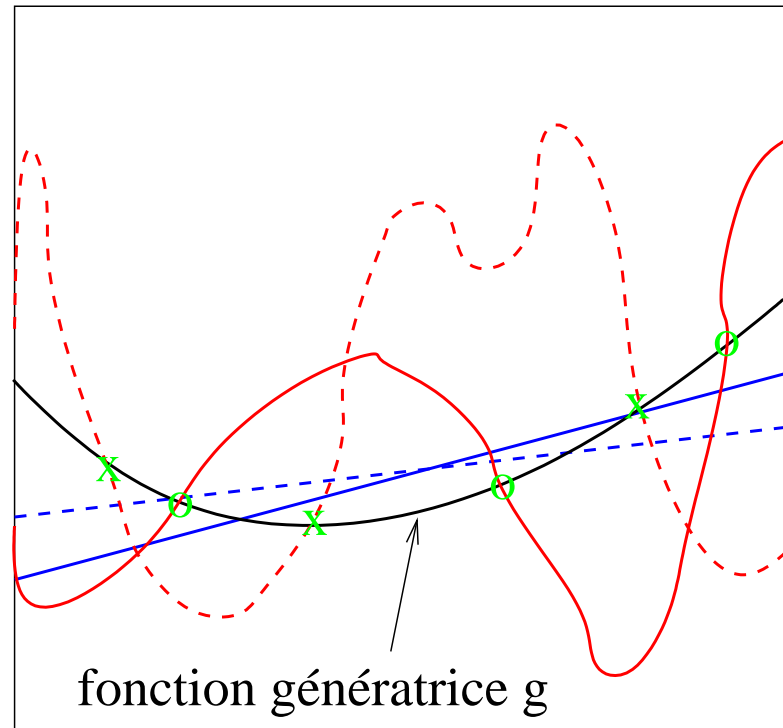
complexité : taille de l'espace des  $f(;w)$ .



- faible complexité
- forte complexité

régularisation : stabilisation de l'espace des solutions par introduction d'a priori.

variance : mesure quadratique de la dépendance de  $f(;w^*)$  à la base expérimentale.



- faible variance
- forte variance

# Contrôle de complexité

---

- Contrôle de la structure de  $f(; w)$ . Expl. : choix du degré d'un polynôme.
- Régularisation : ajouter un terme indépendant de  $x$  dans l'identification du système,

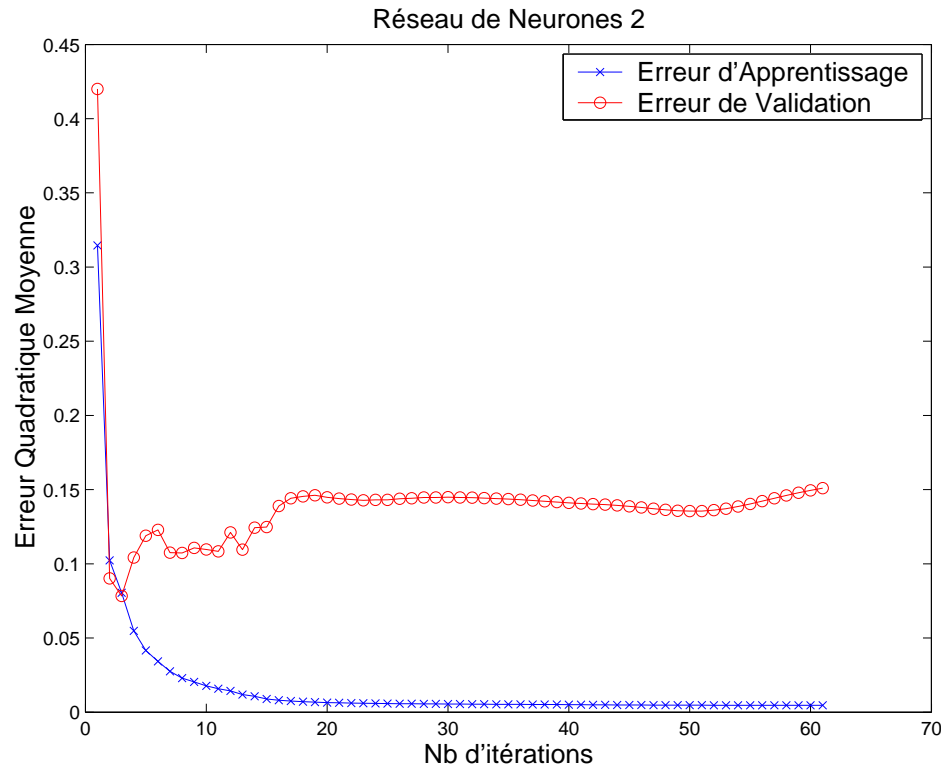
$$\min_w \text{distance}(m, f(x; w)) + \nu R(f(; w)) .$$

- Utilisation d'une erreur de validation : les données sont séparées en un ensemble d'apprentissage  $\mathcal{A}$  qui sert à régler  $w$ , et un ensemble de validation  $\mathcal{V}$  qui sert à calculer l'erreur. But : s'affranchir du biais introduit par la taille finie des données, "ne pas apprendre le bruit".
- + Validation croisée.

# Contrôle de complexité (2)

---

Exemple d'utilisation d'erreur de validation : arrêt prématuré dans les réseaux de neurones.



⇒ le réseau de la troisième itération est utilisé.

# Choix de l'espace de recherche

---

1. Analyse discriminante.
2. Sélection de variables.
3. (Choix de l'intervalle de traitement  $D$ ).

# Analyse discriminante

## (Choix de l'espace de recherche)

---

Recherche de  $p$ ,  $p < d$ , nouvelles variables, combinaisons linéaires des  $x$ , qui séparent en projection les  $k$  classes observées.

Soient,  $W$  la mat. de cov. intraclasse,

$$W = \frac{1}{n} \sum_{i=1}^k \left( \sum_{x_j \in \mathcal{M}_i} (x_j - g_j)(x_j - g_j)^T \right) ,$$

$B$  la mat. de cov. interclasses,  $B = \sum_{i=1}^k \frac{n_i}{n} (g_i - g)(g_i - g)^T$  ,

et  $T$  la mat. de cov. totale,  $T = \frac{1}{n} \sum_{j=1}^n (x_j - g_j)(x_j - g_j)^T = B + W$  .

$\Rightarrow$  Trouver  $p$  axes  $a$  qui maximisent, après projection, l'inertie interclasse,  $a^T B a$  et minimisent l'inertie intraclasse,  $a^T W a$ .

$$\Rightarrow \max_a \frac{a' M B M a}{a' M T M a}$$

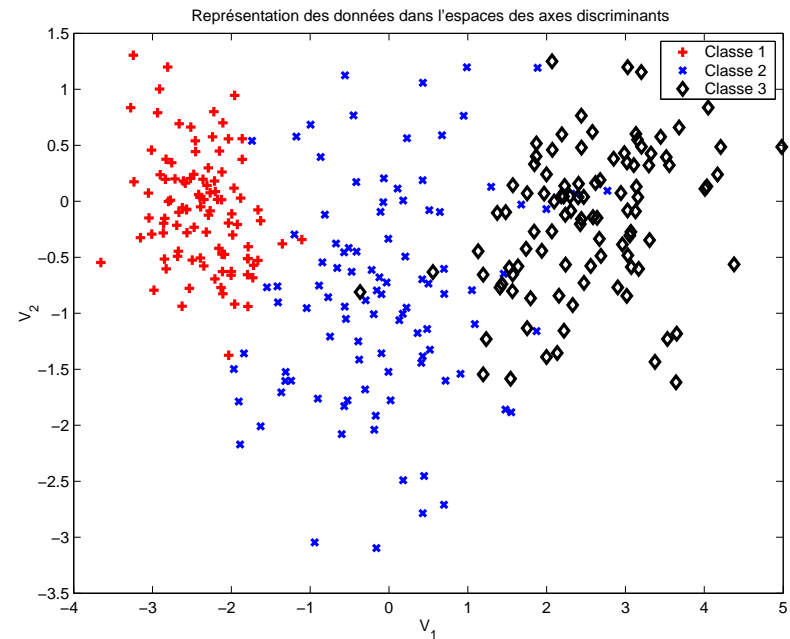
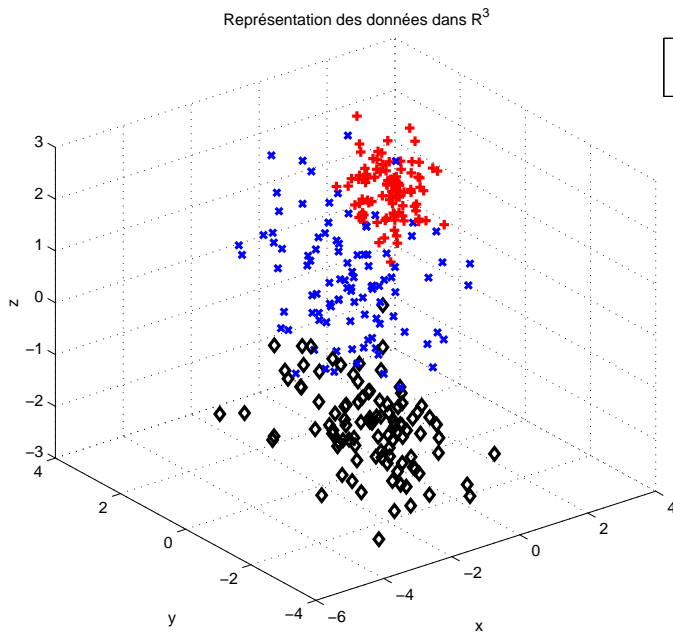
où  $M$  est une métrique ( $= T^{-1}$ ).

# Analyse discriminante (2)

## (Choix de l'espace de recherche)

Les axes discriminants  $a$  sont les vecteurs propres associés aux  $p$  plus grandes valeurs propres  $\lambda$  de

$$BT^{-1}a = \lambda a$$





# Sélection de variables

## (Choix de l'espace de recherche)

---

- Pour connaître les variables les plus pertinentes et réduire la complexité des classeurs.
- L'évaluation d'une sélection par apprentissage est trop coûteuse. On minimise le  $\Lambda$  de Wilks,  $\Lambda = |W|/|T|$ .
- 2 méthodes sont comparées.

### Heuristique d'énumération :

1. Sélectionner la variable parmi les  $x$  dont la suppression induit la plus grande diminution de  $\Lambda$ .
2. La supprimer des  $x$ , recommencer si  $\Lambda >$  seuil.

solution peut être sous-optimale,  
méthode peu coûteuse.

### Algorithme évolutif :

$$\min_{\text{sous-ensembles de } x} \Lambda .$$

optimisation globale  
coûteuse.

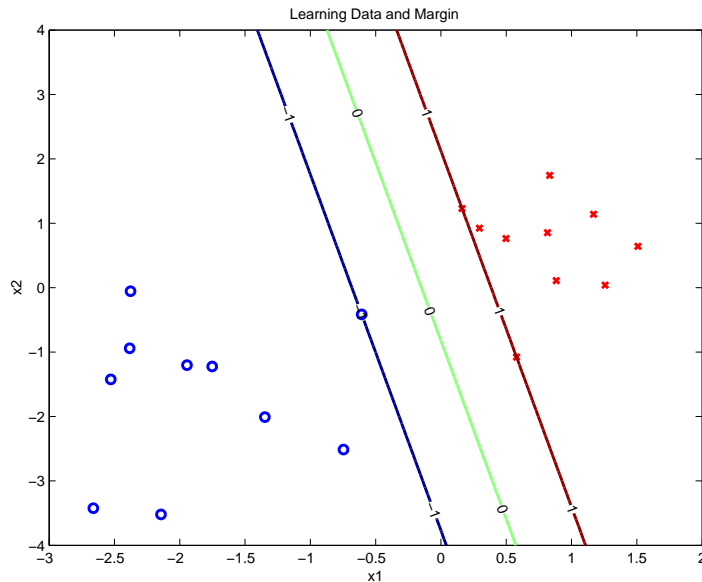
# Méthodes de discrimination

---

Elles sont caractérisées par le critère qu'elles utilisent pour apprendre à classer les exemples.

- Maximisation de marges entre classes : machines à supports vectoriels (MSV).
- Minimisation d'erreur de classement : classeurs Bayésiens.
  - Discrimination par fenêtres de Parzen.
  - Discrimination par régression et seuillage : régression par moindres carrés (linéaire ou neuronale) et  $\epsilon$ -SV régression.

# Classement par Marge maximale : Machines à Supports Vectoriels



Les MSV créent une frontière de décision qui maximise les distances entre la frontière et les points les plus proches de chaque classe.

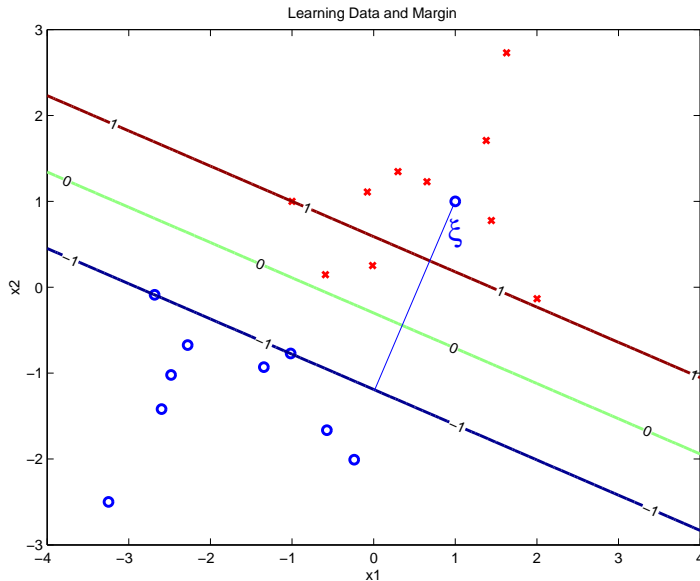
- Frontière de décision linéaire :  
 $f(x) = \langle w, x \rangle + b$

- Maximisation de la marge  $M = \frac{2}{\|w\|}$
- Apprentissage :  $y_i = \pm 1$ ,  
$$\begin{cases} \min_w \frac{1}{2} \|w\|^2 \\ y_i f(x_i) \geq 1 \quad , \quad i = 1, \text{card } \mathcal{A} \end{cases}$$

# MSV, cas non séparable

Lorsque les classes ne sont pas séparables, on introduit des variables d'écart  $\xi_i$  et un paramètre de régularisation  $C$ .

Apprentissage :



$$\begin{cases} \min_{w, \xi_i} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{\text{card } \mathcal{A}} \xi_i \\ y_i f(x_i) \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases}$$

Un problème quadratique, résolu sous sa forme duale (de Wolfe) par un algorithme QP avec stratégie de contraintes actives.

# Extensions aux MSV non-linéaires

- Transformation de l'espace des données,  $x \longrightarrow \Phi(x)$ , puis on applique l'algorithme linéaire aux données transformées.
- Grace à la forme duale, les données apparaissent que sous la forme de produits scalaires,  $\langle \Phi(x), \Phi(x') \rangle = K(x, x')$ ,  $K$  noyau (condition d'existence par Mercer).
- noyau Gaussien de largeur  $\sigma$ ,  $K(x, x') = e^{-\frac{\|x-x'\|^2}{2\sigma^2}}$ .
- noyau polynômial de degré  $o$ ,  $K(x, x') = (\langle x, x' \rangle + 1)^o$ .
- Contrôle de complexité : C, noyau ( $\sigma$  ou  $o$ ).

# Minimisation d'erreur de classement : méthodes Bayésiennes

---

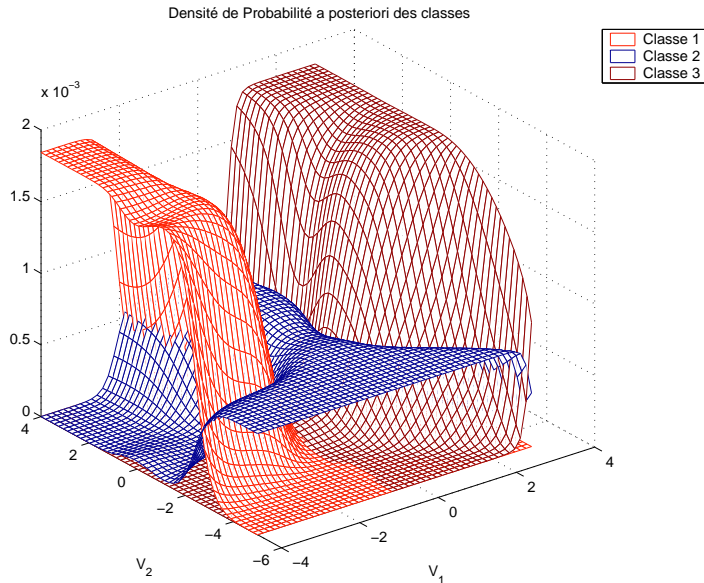
$$x \in \mathcal{M}_i \quad \text{si} \quad P(\mathcal{M}_i|x) > P(\mathcal{M}_j|x) \quad , \quad \forall j \neq i \quad ,$$

ce qui est implique,

$$x \in \mathcal{M}_i \quad \text{si} \quad p(x|\mathcal{M}_i)P(\mathcal{M}_i) > p(x|\mathcal{M}_j)P(\mathcal{M}_j) \quad , \quad \forall j \neq i \quad .$$

La discrimination par fenêtres de Parzen et le seuillage après régression sont des méthodes Bayésiennes.

# Discrimination par fenêtres de Parzen



Une méthode d'histogrammes lissés dont les barres sont centrées sur les données.

$P(\mathcal{M}_k)$  = la fréquence d'apparition de la classe  $\mathcal{M}_k$  dans  $\mathcal{A}$ .

$$p(x|\mathcal{M}_k) = \frac{1}{n_k} \sum_{i \in \{\mathcal{A} \cap \mathcal{M}_k\}} \frac{1}{h^d} \psi \left( \frac{x - x_i}{h} \right) .$$

$\psi$  fenêtre Gaussienne,  $h$  paramètre d'étalement.

# Discrimination par régression et seuillage

---

La masse du véhicule est une variable fonction de  $x$ , scalaire, continue, et physiquement pertinente !

$$\boxed{x \xrightarrow{\text{régression}} \hat{m}(x) \xrightarrow{\text{seuillage}} \text{classe de masse.}}$$

Les seuils sont ajustés par minimisation du taux d'erreur sur la base d'apprentissage  $\mathcal{A}$  (énumération de seuils sur une grille).

Différentes méthodes en fonction de la construction de  $\hat{m}(x)$  : moindres carrés régularisés ou  $\epsilon$ -SV régression.



# Régressions par moindres carrés régularisés

---

$f(; w)$  est un réseau de neurones ou une fonction linéaire.

L'estimation de la masse est  $\hat{m}(x) = f(x; w^*)$  où  $w^*$  est solution de

$$\begin{cases} \min_w J_{2\nu}(w) , \\ J_{2\nu}(w) = J_2(w) + \frac{\nu}{2}w^2 , \\ J_2(w) = \frac{1}{2 \text{card } \mathcal{A}} \sum_{k \in \mathcal{A}} (f(x^{(k)}; w) - m^{(k)})^2 . \end{cases}$$

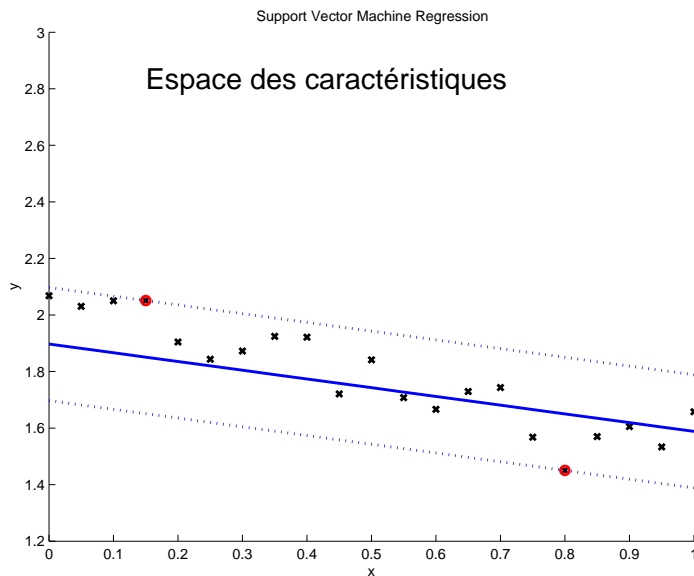
Puisque  $f$  peut être non linéaire (réseau de neurones), on effectue la minimisation avec l'algorithme de Levenberg-Marquardt.

Contrôle de complexité :  $\nu$ , contrôle de la structure de  $f$  (linéaire ou neuronale, combien de neurones dans la couche cachée), arrêt prématuré.

# Régressions par Machines à Supports Vectoriels

## Vectoriels

Les données doivent au plus être à  $\epsilon$  de la fonction  $f(x) = \langle w, x \rangle + b$ , qui doit être la plus plate possible.



$$\left\{ \begin{array}{l} \min_{w,b,\xi_i,\xi_i^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{\text{card } \mathcal{L}} (\xi_i + \xi_i^*), \\ \text{sous les contraintes,} \\ y^{(i)} - \langle w, x^{(i)} \rangle - b \leq \epsilon + \xi_i, \\ \langle w, x^{(i)} \rangle + b - y^{(i)} \leq \epsilon + \xi_i^*, \\ \xi_i, \xi_i^* \geq 0, \\ i = 1, \text{card } \mathcal{L}. \end{array} \right.$$

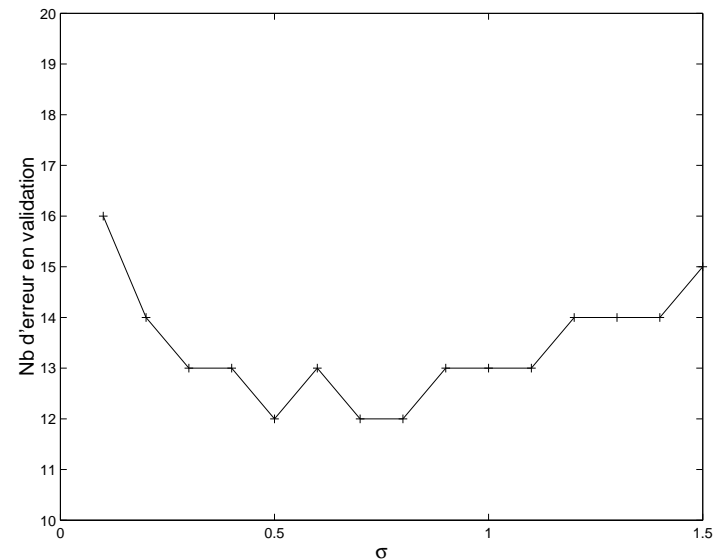
Le problème est équivalent aux MSV en classement (même techniques pour le passage au non-linéaire, mêmes algorithmes).

Contrôle de complexité :  $C$ , noyau ( $\sigma$  ou  $o$ ),  $\epsilon$ .

# Expériences numériques : contrôle de complexité et calcul des erreurs

- Les signaux temporels possèdent une certaine corrélation  $\Rightarrow$  mesure réaliste de l'erreur par validation croisée : plusieurs bases  $\mathcal{A}_k$  et  $\mathcal{V}_k$ , qui respectent la chronologie.
- La plupart des paramètres exerçant un contrôle sur la complexité des méthodes sont ajustés par minimisation de l'erreur de validation croisée. Expl. du  $h$  des fenêtres de Parzen :

1. Pour tous les  $i$ , choisir  $h_i$ .
  - 1.1 Pour tous les  $k$ , apprentissage sur  $\mathcal{A}_k$ , calcul de l'erreur  $E_k$  sur  $\mathcal{V}_k$ .
  - 1.2 Erreur  $E(h_i)$  est la moyenne des  $E_k$ .
2.  $h = \arg \min_i E(h_i)$ .
3. Apprentissage sur  $\mathcal{A} \cup \mathcal{V}$ . 4. Calcul de l'erreur sur  $\mathcal{T}, \mathcal{T} \cap \{\mathcal{A} \cup \mathcal{V}\} = \emptyset$ .



# Variables explicatives pertinentes

	Unité									
Dizaine	0	1	2	3	4	5	6	7	8	9
0		CdgV $M_2$	Vitesse $M_3$	Vitesse $M_2$	Vitesse $M_1$	CdgV $a_{min}$	Vitesse $M_4$	PivArg $M_1$	PivArg $skew$	CdgV $M_3$
1	PivAvg $skew$	Vitesse $moy$	CdgV $M_1$	CdgV $M_4$	PivArd $M_2$	PivArd $M_3$	PivArd $M_2$	PivArd $M_3$	CdgL $M_3$	CdgL $M_2$
2	CdgL $M_4$	PivAvg $kurt$	PivArg $kurt$	CdgT $a_{max}$	CdgT $skew$	Vitesse $skew$	PivArd $var$	PivArd $M_1$	PivArd $M_4$	PivArg $M_3$
3	PivArd $M_4$	Vitesse $a_{min}$	CdgT $var$	CdgT $E$	CdgT $M_2$	PivArd $E$	Vitesse $E$	PivAvg $a_{min}$	PivAvg $E$	CdgV $E$
4	Vitesse $kurt$	CdgV $a_{max}$	PivArg $a_{max}$	PivArd $skew$	PivArd $skew$	PivAvg $M_2$	PivAvg $M_3$	PivAvg $M_4$	PivArg $M_4$	PivArg $a_{min}$
5	PivArd $M_1$	CdgL $E$	CdgL $var$	CdgT $M_3$	CdgT $M_4$	CdgV $kurt$	CdgT $a_{min}$	CdgL $M_1$	Vitesse $a_{max}$	...

Sélection par heuristique énumérative, 55 variables pour le seuil de  $\Lambda$  choisis,

[CdgV et Vitesse] /  $M_i$  sont les variables prépondérantes.

L'algorithme évolutionnaire trouve une solution avec 49 variables.

# Comparaisons de méthodes

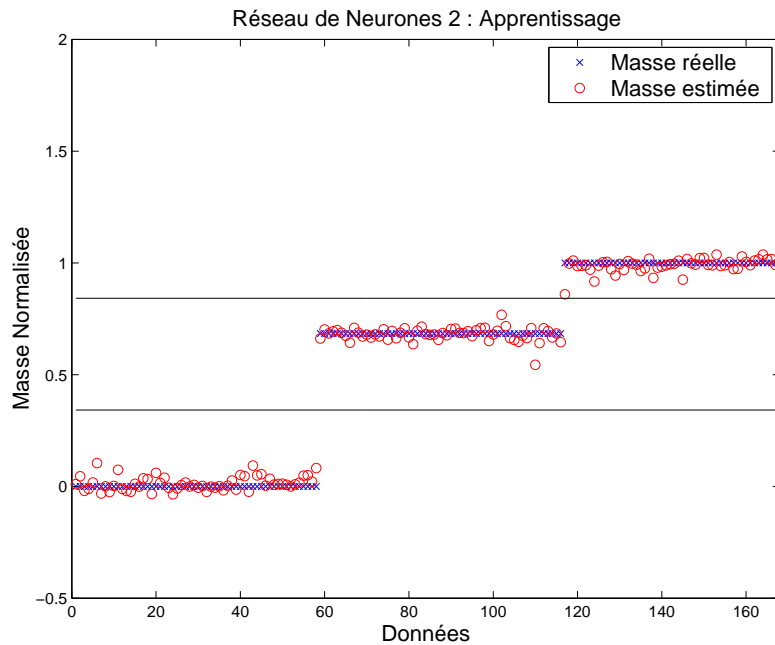
---

5 méthodes sont comparées :

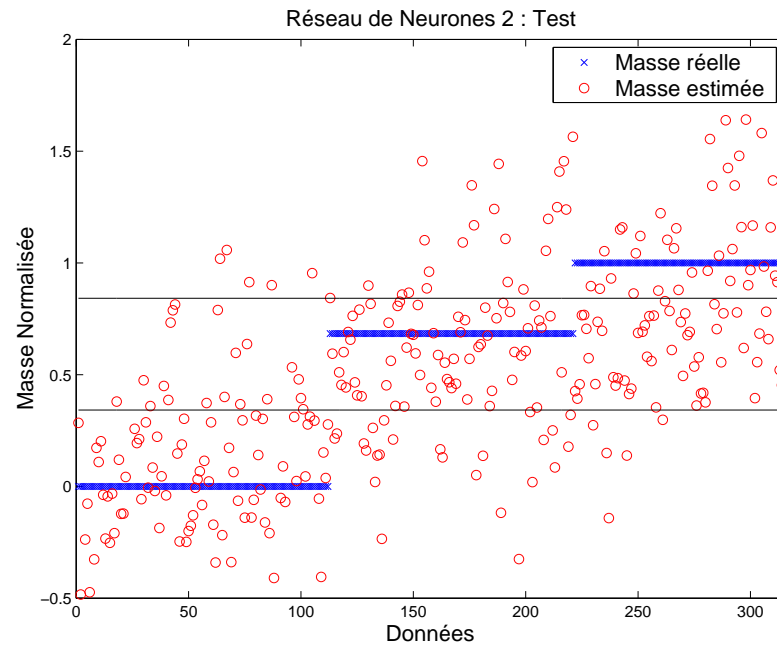
- Analyse discriminante + fenêtres de Parzen.
- Analyse discriminante + MSV classement (noyau Gaussien).
- MSV en classement (noyau Gaussien).
- Régression neuronale + seuillage. S1 = 1 neurone linéaire, S2 = 7 neurones sigmoïdaux + 1 neurone linéaire, S3 = 8 neurones sigmoïdaux.
- MSV en régression (noyau polynomial) + seuillage.

# Exemples de frontières de décisions

Régression neuronale (S2) + seuillage. 81 variables d'entrées. Taux de classification correcte 51.3%



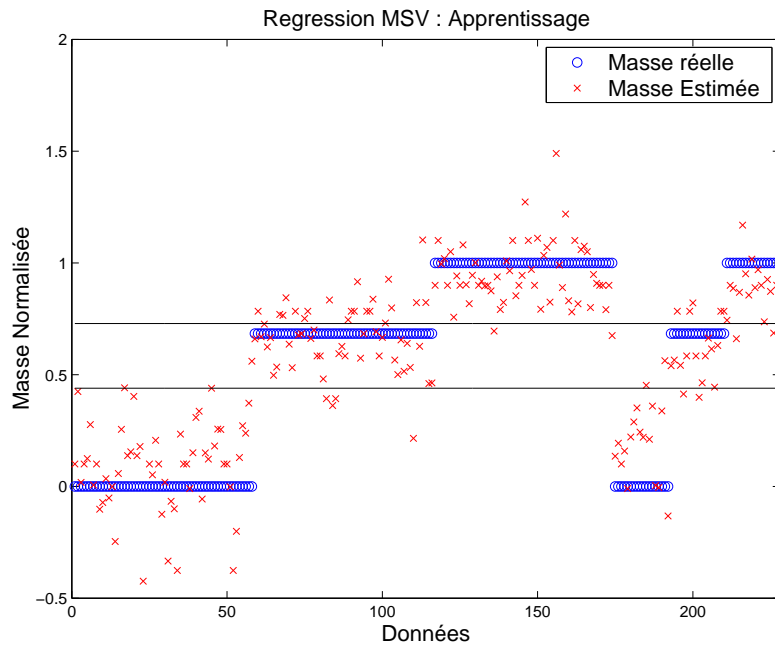
apprentissage



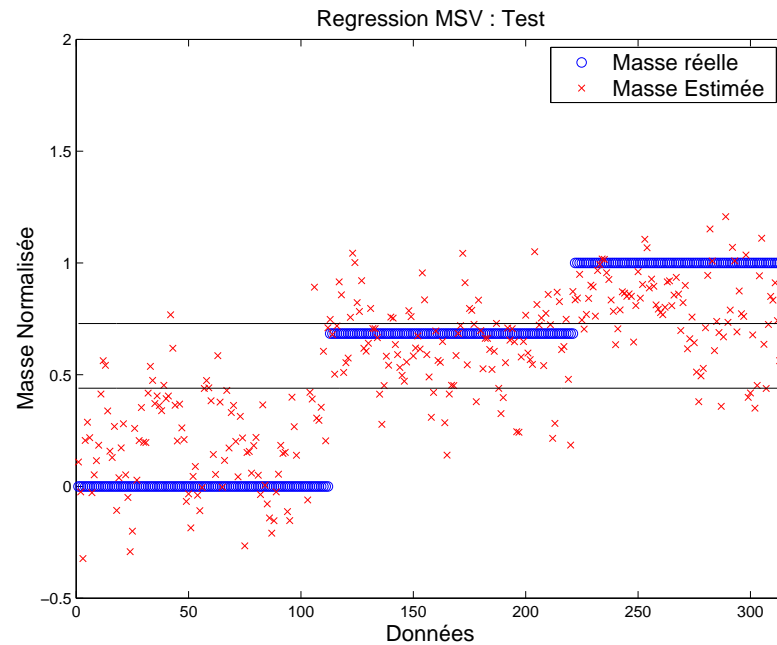
test

# Exemples de frontières de décisions (2)

Régression MSV (noyau polynomial) + seuillage. 81 variables d'entrées. Taux de classification correcte 68.6%.

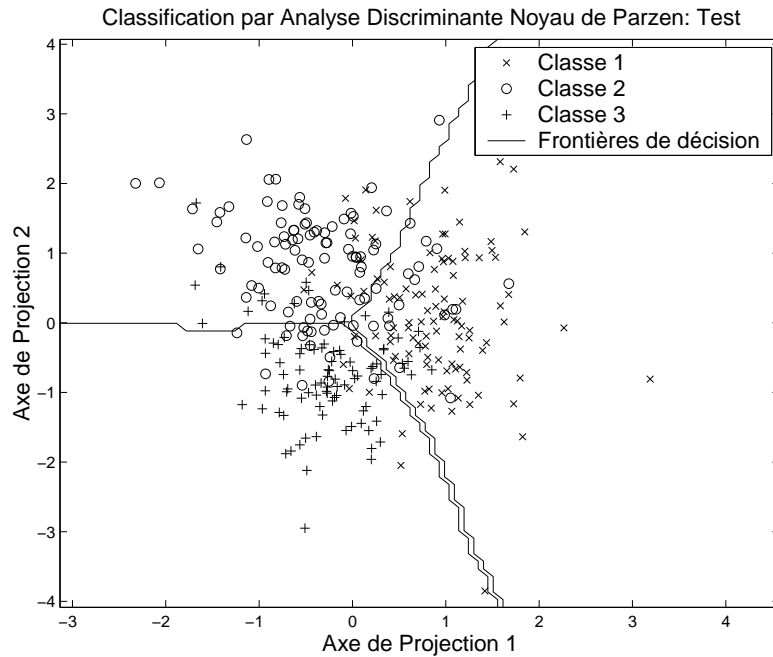


apprentissage

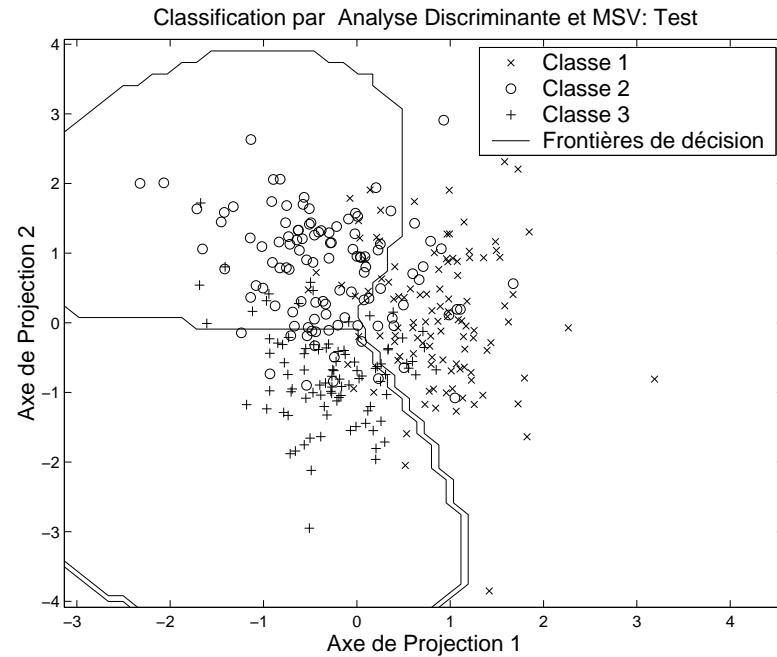


test

# Exemples de frontières de décisions (3)



Analyse discriminante +  
fenêtres de Parzen, 81  
variables d'entrées, taux de  
classification correcte 74.7%.



Analyse discriminante + SVM  
Gaussien, 81 variables  
d'entrées, taux de  
classification correcte 74.6%.



## Comparaison de méthodes (2)

---

Méthodes	B. inf.	moy.	B. sup.
Analyse Discriminante et Parzen	79.25	79.78	80.31
Analyse Discriminante et MSV	80.11	80.61	81.11
MSV à noyau gaussien	49.19	49.63	50.06
Réseaux de neurones 1	63.58	64.16	64.74
Réseaux de neurones 2	52.72	53.38	54.05
Réseaux de neurones 3	50.94	51.90	52.86
MSV à noyau polynomial	71.50	71.78	72.07

Pourcentage de succès de classement pour différentes stratégies en utilisant les 49 variables sélectionnées par l'algorithme évolutionnaire. Les bornes sont l'intervalle de confiance à 80%.

# Conclusions

---

- La meilleure méthode : analyse discriminante + MSV à noyau Gaussien en classement, avec plus de 80% de succès en classement.
- Importance du pré-traitement des variables par analyse discriminante. Cet effet ne s'explique pas seulement par la réduction de dimension des variables d'entrées (car AD fait mieux que la sélection de variables), mais par la prise en compte conjointe des dispersions des données.
- Parmi la régression + seuillage, la régression par MSV donne de bons résultats (71.8% de succès), meilleurs que la régression linéaire ou neuronale.
- Les signaux les plus importants sont l'accélération verticale au cdg et la vitesse.
- Les traitements les plus importants sont les moments spectraux relatifs.