

Modèle de régression à processus logistique caché pour le diagnostic par suivi de point de fonctionnement des mécanismes d'aiguillages

Faïcel Chamroukhi

Direction de la thèse : Gérard Govaert (UTC), Patrice Aknin(INRETS)

Encadrement : Allou Samé (INRETS)

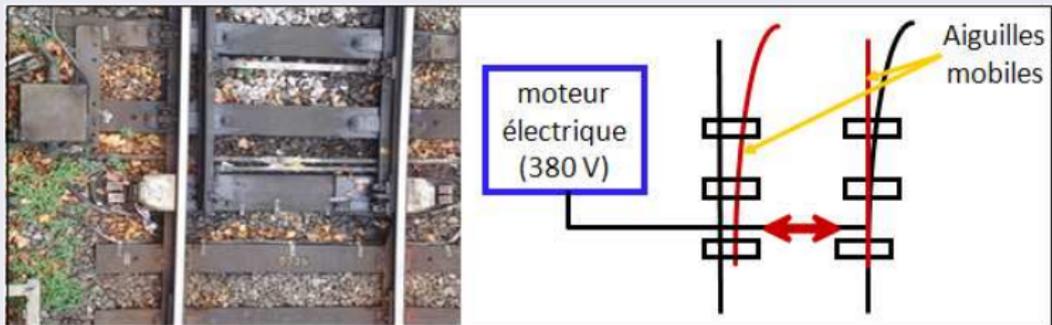


- 1 Introduction
 - Contexte
 - Objectifs visés
 - Approche proposée
- 2 Modélisation probabiliste des signaux
 - Modèle de régression à processus logistique caché
 - Le processus Logistique proposé
 - Estimation des paramètres du modèle
- 3 Suivit de point de fonctionnement
 - Modélisation de la dynamique d'évolution de l'état du système
- 4 Experimentations
- 5 Conclusion
- 6 Perspectives

Contexte

- Collaboration avec la SNCF
- Diagnostic par suivi de point de fonctionnement du mécanisme d'aiguillage
- Application des méthodes développées au circuit de voie

Mcanisme d'aiguillage

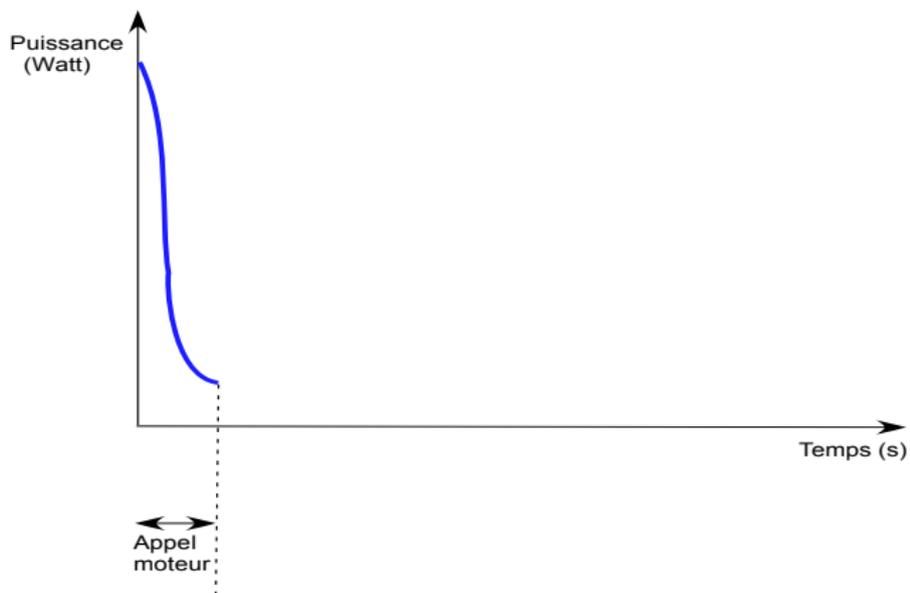


Objectifs visés

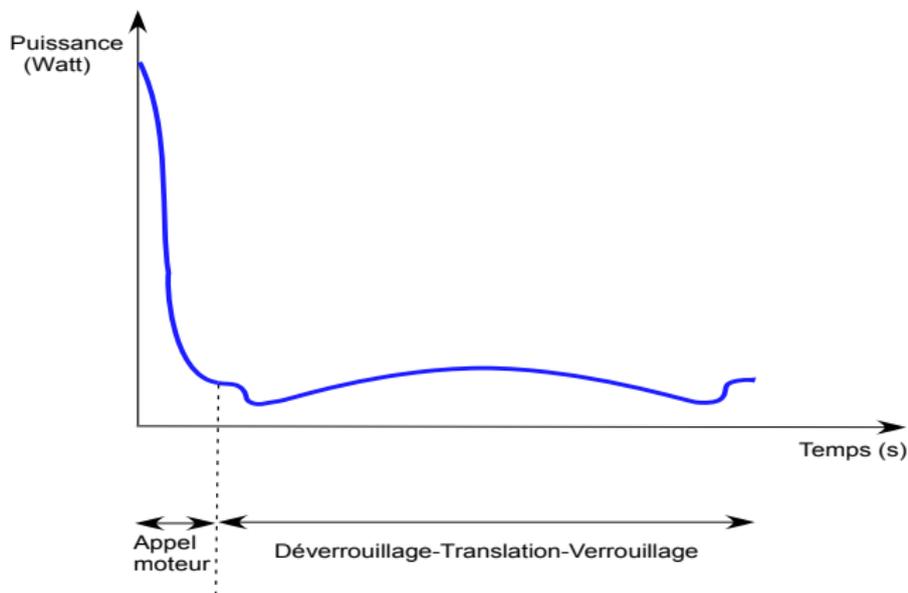
- Suivi du point de fonctionnement :
 - Estimer la dynamique d'évolution de l'état de fonctionnement du système
- Détecter précisément le type défaut (défaut mécanique, défaut électrique, défaut de graissage,...) afin de déployer l'équipe concerne

Donnes disponibles pour le diagnostic : Puissance consommée par le moteur durant la manœuvre

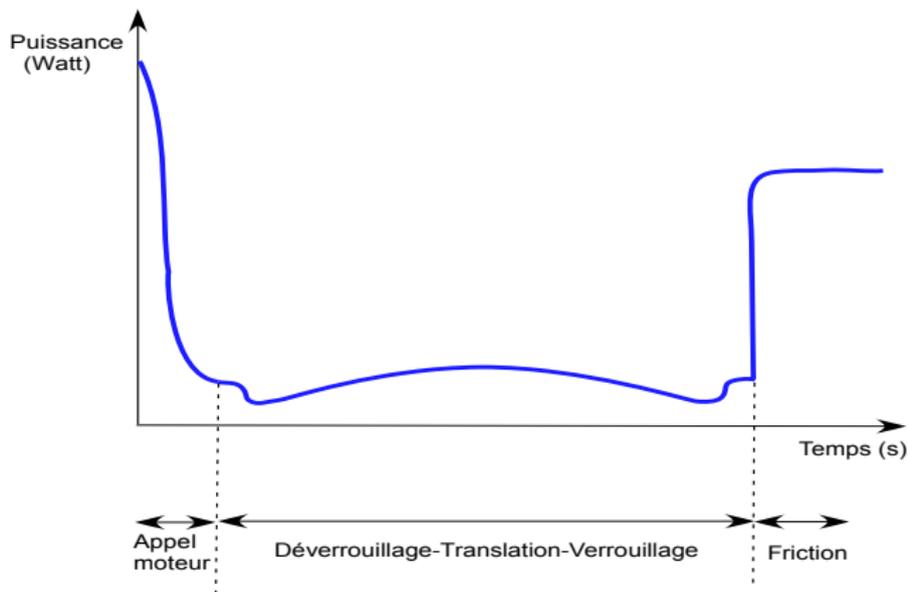
Donnes disponibles pour le diagnostic : Puissance consommée par le moteur durant la manœuvre



Donnes disponibles pour le diagnostic : Puissance consommée par le moteur durant la manœuvre

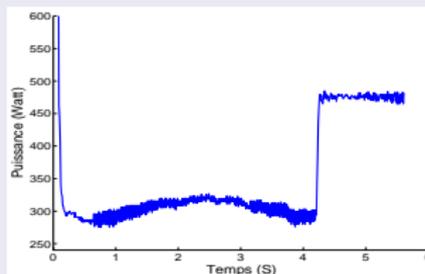


Donnes disponibles pour le diagnostic : Puissance consommée par le moteur durant la manœuvre

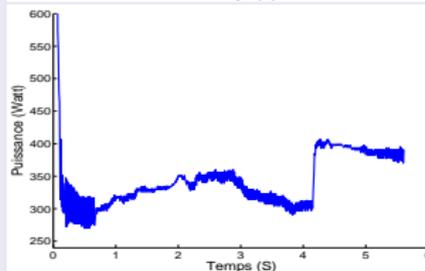


Exemples d'états de fonctionnement de l'aiguillage

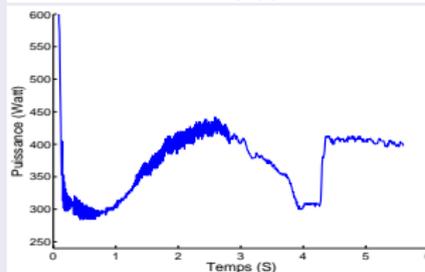
Manœuvre sans défaut



Manœuvre avec défaut mineur

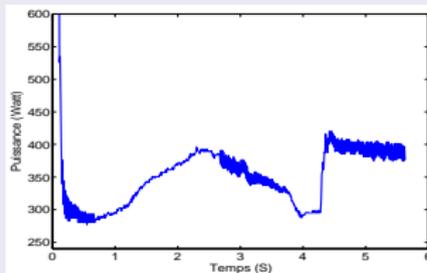


Manœuvre avec défaut critique

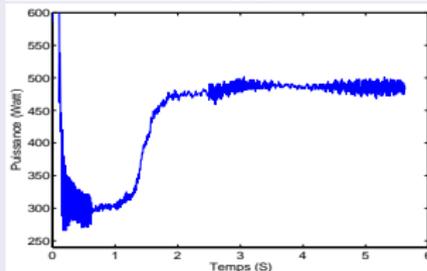


Variabilité des signatures des défauts

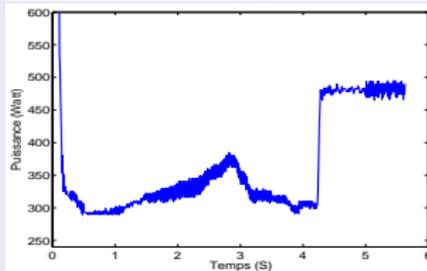
Défaut de graissage



Friction anticipée



Mauvais réglage de l'attaque multiple



Approche proposée

- 1 ère étape : **Caractérisation** de chaque signal dans un espace de représentation spécifique
 - Estimation des paramètres d'un modèle probabiliste de signal
 - Avantages :
 - apte à prendre en compte les différentes phases des signaux étudiés
 - flexibilité
 - possibilité de simuler facilement
- 2 ème tape : Développement d'une **méthode de suivi temporel** s'appuyant sur les paramètres estimés.

- Chaque signal est représenté par une séquence temporelle de n observations réelles :

$\mathbf{x} = (x_1, \dots, x_n)$ où x_i est observé à l'instant t_i .

- Échantillon : $\{(t_1, x_1), \dots, (t_n, x_n)\}$
 - t_i : le temps
 - x_i : variable aléatoire réelle représentant le signal l'instant t_i
- \mathbf{x} est supposé être généré par un modèle de régression à processus logistique caché $\mathbf{z} = (z_1, \dots, z_n)$ avec :
 - z_i variable aléatoire discrète prend ses valeurs dans $\{1, \dots, K\}$
 - z_i le label du modèle de régression associé à x_i (classe de x_i).

Le processus Logistique

Etant donnés les instants (t_1, \dots, t_n) :

- $z_i \sim \mathcal{M}(1, \pi_{i1}(\mathbf{w}), \dots, \pi_{iK}(\mathbf{w}))$

$$\pi_{ik}(\mathbf{w}) = p(z_i = k; \mathbf{w}) = \frac{\exp(\mathbf{w}_k^T \mathbf{v}_i)}{\sum_{\ell=1}^K \exp(\mathbf{w}_\ell^T \mathbf{v}_i)}, \quad (1)$$

Où $\mathbf{w} = (\mathbf{w}_1, \dots, \mathbf{w}_K)$ et $\mathbf{w}_k = (\mathbf{w}_{k0}, \dots, \mathbf{w}_{kq})^T$ est le vecteur $(q+1)$ -dimensionnel des coefficients associés à $\mathbf{v}_i = (1, t_i, \dots, (t_i)^q)^T$.

- La distribution de \mathbf{z} s'écrit :

$$p(\mathbf{z}; \mathbf{w}) = \prod_{i=1}^n \prod_{k=1}^K \left(\frac{\exp(\mathbf{w}_k^T \mathbf{v}_i)}{\sum_{\ell=1}^K \exp(\mathbf{w}_\ell^T \mathbf{v}_i)} \right)^{z_{ik}} \quad (2)$$

- $z_{ik} = 1$ si $z_i = k$.
- $z_{ik} = 0$ sinon.

Propriétés du processus logistique

- 1 Effet de la dimension q de \mathbf{w}_k sur l'évolution temporelle des probabilités $\pi_{ik}(\mathbf{w})$.
- On considère un exemple de $K = 2$ classes.
- Seule $\pi_{i1}(\mathbf{w}) = \frac{\exp(\mathbf{w}_1^T \mathbf{v}_i)}{1 + \exp(\mathbf{w}_1^T \mathbf{v}_i)}$ est présentée ($\pi_{i2}(\mathbf{w}) = 1 - \pi_{i1}(\mathbf{w})$).

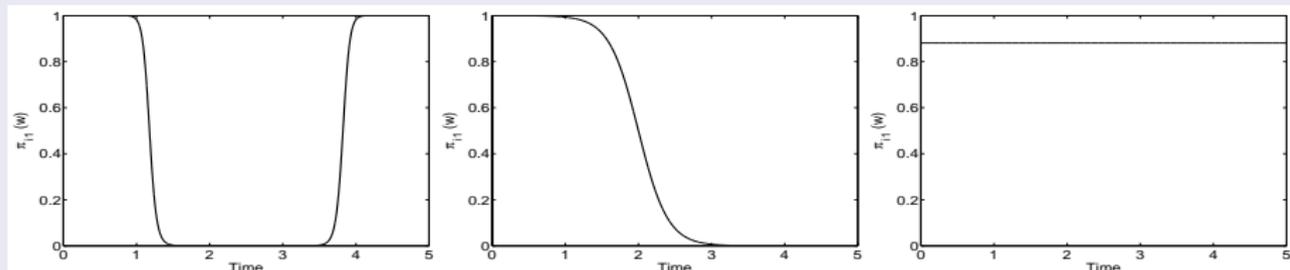


FIGURE: Exemple de probabilités logistiques en fonction du temps pour différentes valeurs de q , avec $K = 2$ et (gauche) $q = 2$ et $\mathbf{w}_1 = (-10, -20, -4)^T$, (milieu) $q = 1$ et $\mathbf{w}_1 = (10, -5)^T$, (droite) $q = 0$ et $\mathbf{w}_1 = 2$.

⇒ Pour avoir des classes (segments) connexes, q doit être fixé à 1.

Propriétés du processus logistique

2 Evolution des probabilités $\pi_{ik}(\mathbf{w})$ au cours du temps en fonction de \mathbf{w}_k pour une dimension q de \mathbf{w}_k fixe (ici $q = 1$).

- Paramétrisation de $\mathbf{w}_k = (\mathbf{w}_{k0}, \mathbf{w}_{k1})^T$ par $\mathbf{w}_k = \lambda_k(\gamma_k, 1)^T$.

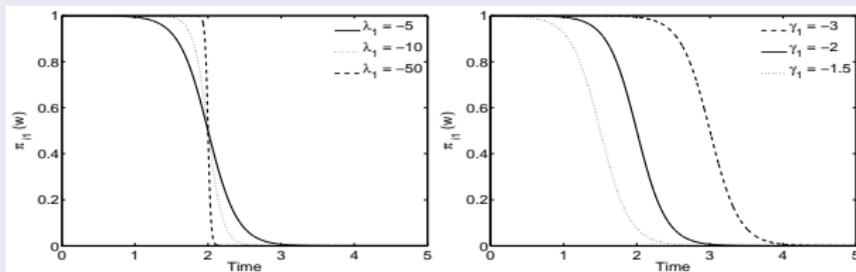


FIGURE: Exemple de probabilités logistiques en fonction du temps pour différentes configurations de \mathbf{w} , avec $q = 1$, $K = 2$ et (gauche) $\gamma_1 = -2$ et $\lambda_1 = [-5, -10, -50]$, (droite) $\lambda_1 = -5$ et $\gamma_1 = [-3, -2, -1.5]$.

$\Rightarrow \lambda_k$ contrôle **la qualité des transitions entre les classes** (plus $|\lambda_k|$ est élevée plus la transition entre les z_i est abrupte).

\Rightarrow Le paramètre γ_k contrôle **le temps de changement**.

Modèle de régression

- Signal

- Échantillon : $\{(t_1, x_1), \dots, (t_n, x_n)\}$
- t_i : le temps
- x_i : variable aléatoire réelle représentant le signal l'instant t_i

- Modèle proposé

$$\forall i = 1, \dots, n, \quad \begin{cases} x_i = \beta_{z_i}^T \mathbf{r}_i + \varepsilon_i \\ \varepsilon_i \sim \mathcal{N}(0, \sigma_{z_i}^2), \end{cases} \quad (3)$$

- $z_i \in \{1, \dots, K\}$ le label du sous-modèle de régression de x_i
- $\Rightarrow z_i$ permet de basculer de façon automatique d'un modèle régressif l'autre
- $\mathbf{r}_i = (1, t_i, \dots, t_i^p)$; p ordre de régression
- $\beta_k = (\beta_{k0}, \dots, \beta_{kp})$: coefficients de régression

Loi de probabilité des observations $(x_i, i = 1, \dots, n)$

En faisant intervenir la **variable latente discrète** z représentant la classe de chaque observation x_i , on montre que :

$$\begin{aligned}
 p(x_i; \theta) &= \sum_{k=1}^K p(x_i, z_i = k; \theta) \\
 &= \sum_{k=1}^K p(z_i = k; \mathbf{w}) p(x_i | z_i = k; \theta) \\
 &= \sum_{k=1}^K \pi_{ik}(\mathbf{w}) \mathcal{N}(x_i; \beta_k^T \mathbf{r}_i, \sigma_k^2)
 \end{aligned}$$

⇒ **modèle de mélange de lois** à chaque instant t_i

Paramètres du modèle

$$\theta = \left(\mathbf{w}_1, \dots, \mathbf{w}_K, \beta_1, \dots, \beta_K, \sigma_1^2, \dots, \sigma_K^2 \right)$$

Estimation du vecteur θ des paramètres du modèle

- Méthode du maximum de vraisemblance

- $x_i = \beta_{z_i}^T \mathbf{r}_i + \varepsilon_i$

- $\varepsilon_1, \dots, \varepsilon_n$ sont supposées indépendants cond. aux $z_i \Rightarrow$ Étant donné le temps $\mathbf{t} = (t_1, \dots, t_n)$, les x_i sont indépendants

- Log-vraisemblance

$$\begin{aligned} L(\theta) &= \log P(x_1, \dots, x_n; \theta) \\ &= \sum_{i=1}^n \log \left[\sum_{k=1}^K \pi_{ik}(\mathbf{w}) \mathcal{N}(x_i; \beta_k^T \mathbf{r}_i, \sigma_k^2) \right] \end{aligned}$$

La maximisation directe de la log-vraisemblance ne peut pas être effectuée de manière analytique

- Maximisation par **l'algorithme EM (Expectation Maximization)**
[DLR 77]

Algorithme EM

Maximiser à chaque itération (m) $Q(\theta; \theta^{(m)})$

$$\begin{aligned} Q(\theta; \theta^{(m)}) &= E(L_c(\theta; \mathbf{Y}) | \mathbf{x}, \theta^{(m)}) \\ &= \sum_{i=1}^n \sum_{k=1}^K t_{ik}^{(m)} \log \pi_{ik}(\mathbf{w}) + \sum_{i=1}^n \sum_{k=1}^K t_{ik}^{(m)} \log \mathcal{N}(x_i; \beta_k^T \mathbf{r}_i, \sigma_k^2) \end{aligned}$$

avec :

- $\mathbf{x} = (x_1, \dots, x_n)$: données observées
- $\mathbf{y} = (x_1, \dots, x_n, \mathbf{z}_1, \dots, \mathbf{z}_n)$: données complètes
- $L_c(\theta; \mathbf{y})$: log-vraisemblance complète

⇒ La maximisation de $Q(\theta; \theta^{(m)})$ à chaque itération m fait croître la vraisemblance $L(\theta; \mathbf{x})$

Détails de l'algorithme EM obtenu

Détails de l'algorithme EM obtenu

- Initialisation : $\theta^{(0)}$

Détails de l'algorithme EM obtenu

- Initialisation : $\theta^{(0)}$

Itérer jusqu' la convergence les deux étapes suivantes :

Détails de l'algorithme EM obtenu

- Initialisation : $\theta^{(0)}$

Itérer jusqu' la convergence les deux étapes suivantes :

- 1 Étape E : Espérance (à l'itération m)

Calcul de l'espérance conditionnelle de la vraisemblance complète $Q(\theta; \theta^{(m)})$

⇒ nécessite simplement le calcul des probabilités a posteriori $t_{ik}^{(m)}$, $\forall i$, $\forall k$

$$t_{ik}^{(m)} = \frac{\pi_{ik}(\mathbf{w}^{(m)}) \mathcal{N}(x_i; \beta_k^T \mathbf{r}_i, \sigma_k^2)^{2(m)}}{\sum_{\ell=1}^K \pi_{\ell i}(\mathbf{w}^{(m)}) \mathcal{N}(x_i; \beta_\ell^T \mathbf{r}_i, \sigma_\ell^2)^{2(m)}}$$

Détails de l'algorithme EM obtenu

- Initialisation : $\theta^{(0)}$

Itérer jusqu' la convergence les deux étapes suivantes :

- 1 Étape E : Espérance (à l'itération m)

Calcul de l'espérance conditionnelle de la vraisemblance complète $Q(\theta; \theta^{(m)})$

⇒ nécessite simplement le calcul des probabilités a posteriori $t_{ik}^{(m)}$, $\forall i$, $\forall k$

$$t_{ik}^{(m)} = \frac{\pi_{ik}(\mathbf{w}^{(m)}) \mathcal{N}(x_i; \beta_k^T \mathbf{r}_i, \sigma_k^2)^{2(m)}}{\sum_{\ell=1}^K \pi_{\ell i}(\mathbf{w}^{(m)}) \mathcal{N}(x_i; \beta_\ell^T \mathbf{r}_i, \sigma_\ell^2)^{2(m)}}$$

- 2 Étape M : Maximisation (à l'itération m)

Calcul de $\theta^{(m+1)}$ qui maximise $Q(\theta; \theta^{(m)})$ par rapport au paramètre θ

Détails de l'étape M

Maximisation de $Q(\theta; \theta^{(m)}) = Q_1(\mathbf{w}) + Q_2(\beta, \sigma_1^2, \dots, \sigma_K^2)$:

Détails de l'tape M

Maximisation de $Q(\theta; \theta^{(m)}) = Q_1(\mathbf{w}) + Q_2(\beta, \sigma_1^2, \dots, \sigma_K^2)$:

- Calcul du $\mathbf{w}^{(m+1)}$ qui maximise Q_1 : problème de régression logistique multicalasse pondéré par les $t_{ik}^{(m)}$ (Algorithme IRLS (Iterative Reweighted Least Square [Chen 99, Green 84, Krishnapuram 05])

Détails de l'étape M

Maximisation de $Q(\theta; \theta^{(m)}) = Q_1(\mathbf{w}) + Q_2(\beta, \sigma_1^2, \dots, \sigma_K^2)$:

- Calcul du $\mathbf{w}^{(m+1)}$ qui maximise Q_1 : problème de régression logistique multicalasse pondéré par les $t_{ik}^{(m)}$ (Algorithme IRLS (Iterative Reweighted Least Square [Chen 99, Green 84, Krishnapuram 05])
- La maximisation de Q_2 par rapport à β_k consiste à résoudre K problèmes de régression linéaire pondérés par les $t_{ik}^{(m)}$.

$$(\beta_k^T)^{(m+1)} = \underset{\beta_k}{\operatorname{argmin}} \sum_{i=1}^n t_{ik}^{(m)} (x_i - \beta_k^T \mathbf{r}_i)^2 \quad ; \quad \forall k = 1, \dots, K$$

- La maximisation de Q_2 par rapport à (σ_k^2) donne :

$$(\sigma_k^2)^{(m+1)} = \frac{\sum_{i=1}^n t_{ik}^{(m)} (x_i - (\beta_k^T)^{(m+1)} \mathbf{r}_i)^2}{\sum_{i=1}^n t_{ik}^{(m)}} \quad (\text{solution de l'équation de vraisemblance})$$

Algorithme Iteratively Reweighted Least Squares (IRLS)

- Utilisé pour maximiser $Q_1 = \sum_{i=1}^n \sum_{k=1}^K t_{ik}^{(m)} \log \pi_{ik}(\mathbf{w})$ par rapport à \mathbf{w} .
- Q_1 est la vraisemblance du paramètre \mathbf{w} étant donné le temps et l'espérance du processus caché \mathbf{z} .
- Typiquement un algorithme Newton-Raphson dont la formule de mise à jour est :

$$\mathbf{w}^{(c+1)} = \mathbf{w}^{(c)} - \alpha [H(\mathbf{w}^{(c)})]^{-1} g(\mathbf{w}^{(c)}) \quad (4)$$

où $H(\mathbf{w})$ et $g(\mathbf{w})$ sont respectivement la matrice Hésienne et le gradient de \mathbf{w} , et α le pas d'apprentissage.

- Contrairement à [Krishnapuram 05] (approximation de la hésienne pour accélérer l'algorithme) on utilise la matrice Hésienne pleine [Chen 99] pour une meilleure estimation des paramètres.

Modèle de régression multivariée à processus caché

$\mathbf{X} = \{x_1, \dots, x_N\}$ une séquence de signaux observée sur un interval de temps $\{t_1, \dots, t_N\}$.

- Échantillon : $\{(t_1, x_1), \dots, (t_N, x_N)\}$
- t_j : le temps (inter-signaux)
- x_j : variable aléatoire réelle représentant les paramètres du signal l'instant t_j

\mathbf{X} est supposé être généré par un modèle de regression à processus logistique caché $\mathbf{C} = (c_1, \dots, c_N)$ où :

- c_j variable aléatoire discrete prend ses valeurs dans $\{1, \dots, G\}$
- c_j l'état du système à l'instant t_j .

Modèle de régression

- Séquence de signaux \rightarrow paramétrisation \rightarrow paramètres
- Désormais, on désigne par x_j : variable aléatoire dans R^d représentant les paramètres du signal observé à l'instant t_j
- Modèle proposé :

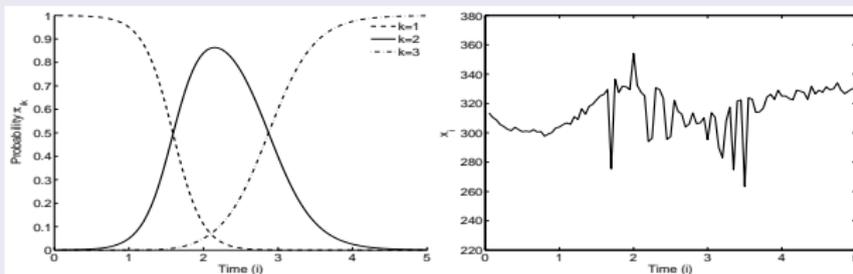
$$\forall j = 1, \dots, N, \quad \begin{cases} x_j = B_{c_j}^T \mathbf{r}_j + \varepsilon_j \\ \varepsilon_j \sim \mathcal{N}(\mathbf{0}, \Sigma_{c_j}), \end{cases} \quad (5)$$

- $c_j \in \{1, \dots, G\}$ variable latente discrète qui représente l'état du système à l'instant t_j .
- $\mathbf{r}_j = (1, t_j, \dots, t_j^p)$; p ordre de régression
- $B_g = (B_{g0}, \dots, B_{gp})$: matrice de dim. $[(p+1) \times d]$ des coefficients de régression multivariée associées à \mathbf{r}_j .

Donnée simulés

- Simuler des données avec $K = 3$ classes, $p = 2$ et $q = 1$ à partir de paramètres θ fixés correspondant à une erreur théorique de 3%.
- Faire varier la taille de l'échantillon n (n varie de 100 à 1000).
- Qualité des estimateurs mesurée par l'erreur de segmentation par rapport la segmentation de référence (simulée).
- Résultats calculés pour 20 échantillons différents.

Exemple de signal simulé ($n = 100$) :



Erreur de classification

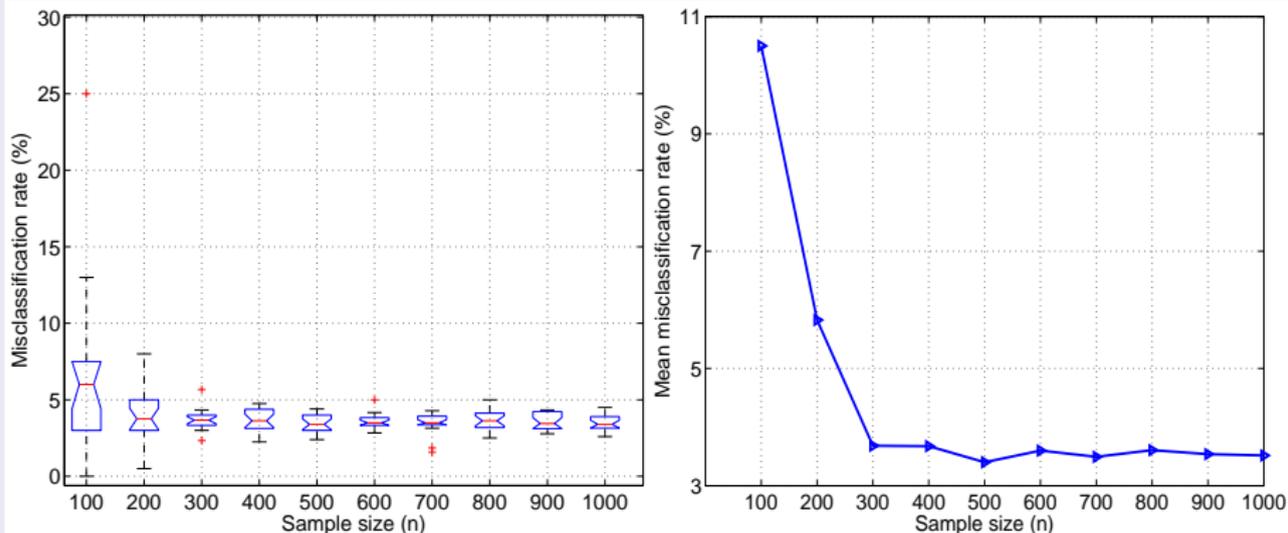


FIGURE: Misclassification rates (left) and mean misclassification rates (right) for different sample sizes n with a theoretical Bayes error around 3%.

Selection de modèle par le critère BIC [schwartz 78]

K	2	3	4	5	6	7	8
p							
0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
2	0	85	5	0	0	0	0
3	0	5	5	0	0	0	0
4	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0

TABLE: Percentage of choosing (K, p) with BIC criterion

Selection de modèle par le critère BIC [schwartz 78]

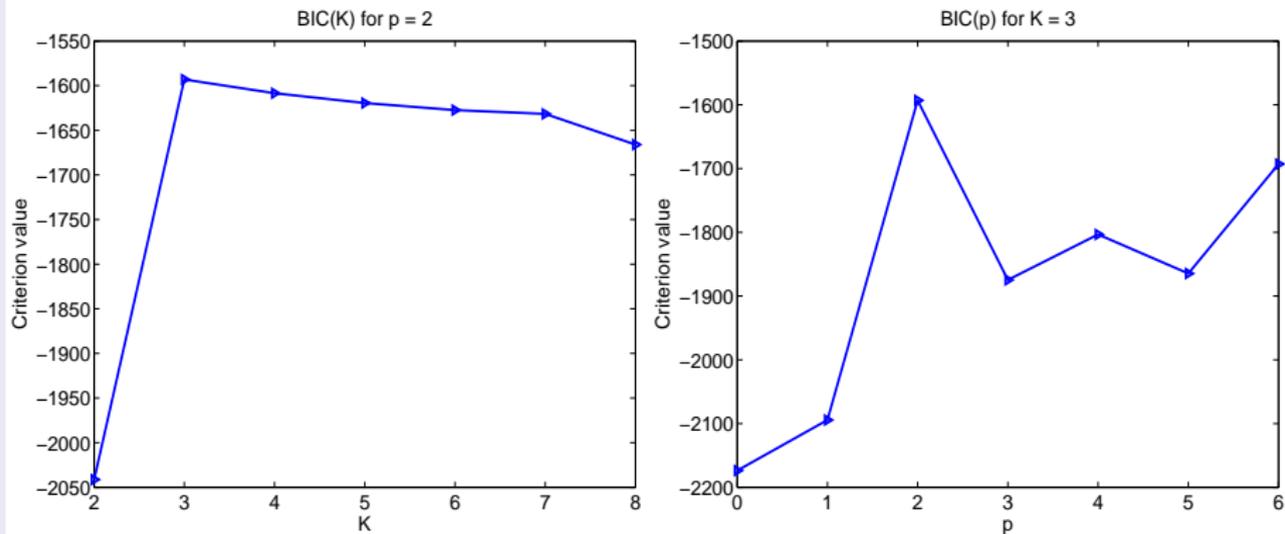
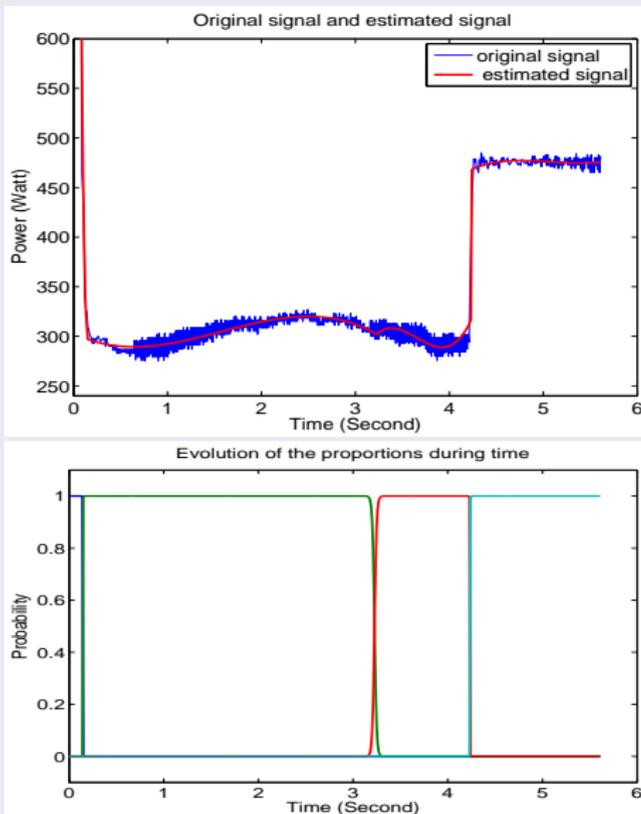


FIGURE: Choosing K and p with BIC for one example of simulated signals

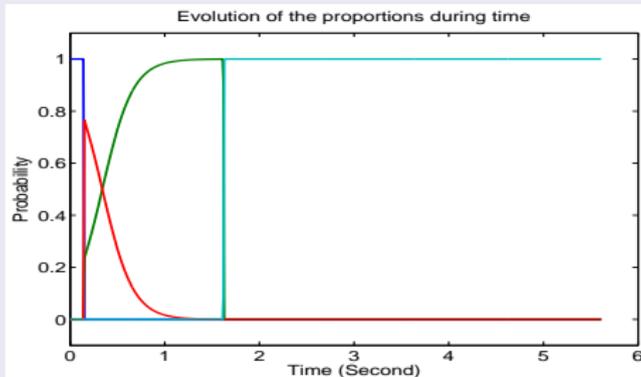
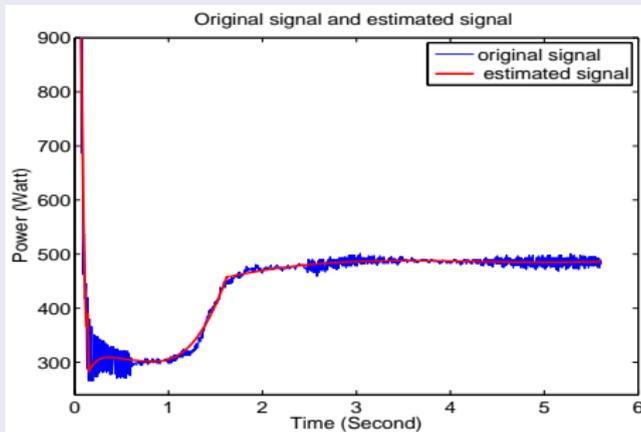
Données réelles

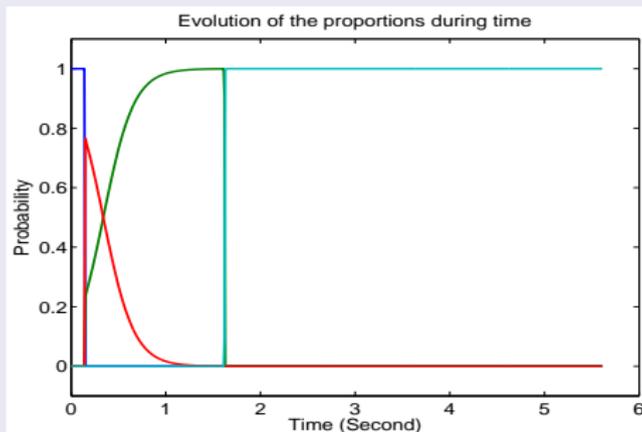
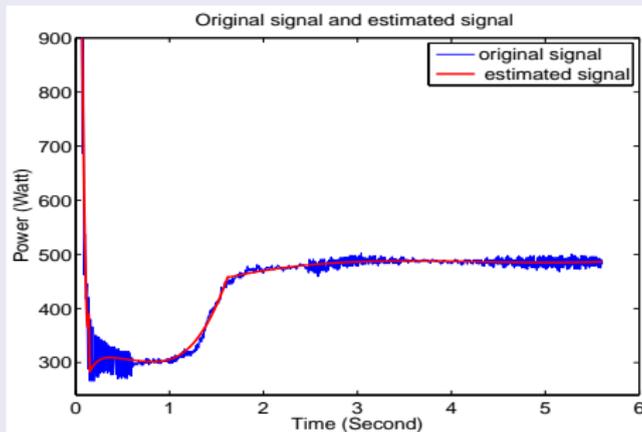
- Signaux réels couvrant un large panel de défauts
- Nombre de composantes du mélange choisi : $K = 4 \simeq$ nombre de phases d'une manœuvre.
- Ordre de régression linéaire $p = 3$ (splines cubiques).
- Dimension du paramètres du logistique $q = 1$.
 \Rightarrow Nombre de paramètres à estimer par signal :
 $(p + q + 3) * K - (q + 1) = 26$ (Vs. 560 pts initiaux).

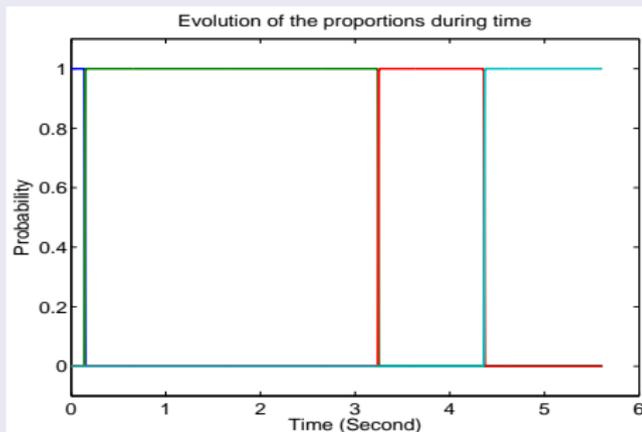
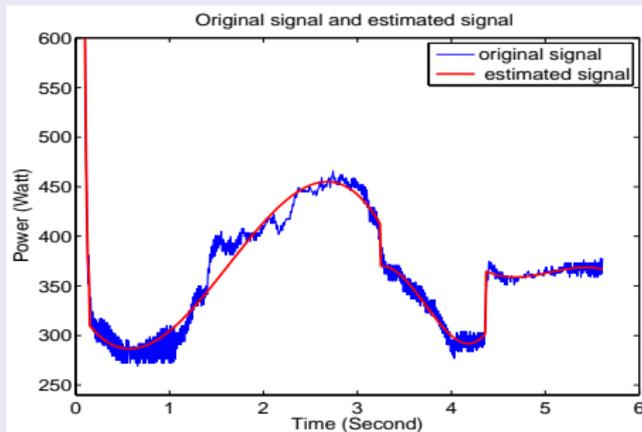
Cas d'un signal sans défaut



Cas d'un signal avec défaut

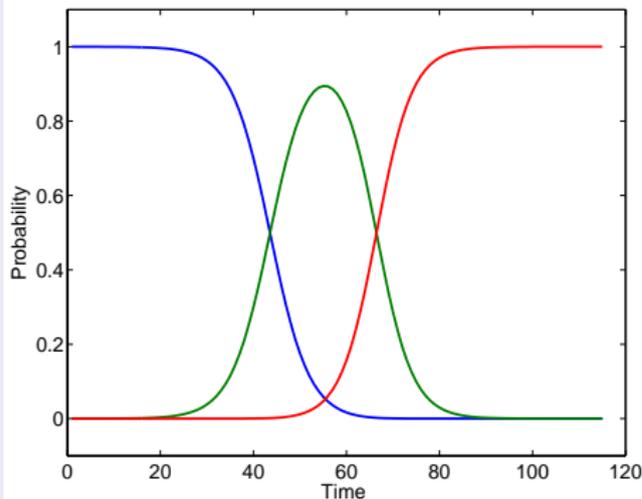




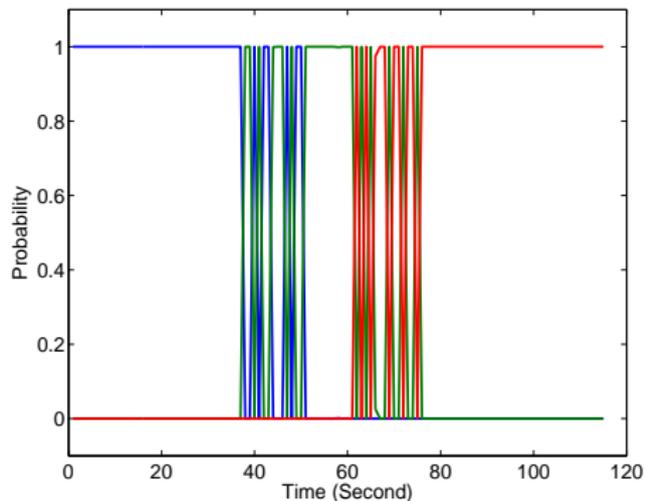


Résultats pour le suivi de point de fonctionnement

Evolution of the state probabilities over time



Evolution of the posterior probabilities during time



Modèle concurrent : La séquence des classes (états) est un processus Markovien (HMM).

- Hidden Markov processes, more commonly known as hidden Markov models (HMMs)
- Modèle à K états (classes) caché(e)s.
- Loi initiale $\pi_1 = (\pi_{11}, \dots, \pi_{1K})^T$, avec $\pi_{1k} = p(z_1 = k)$
- Matrice de transitions A , avec $A_{\ell k} = p(z_i = k | z_{i-1} = \ell)$

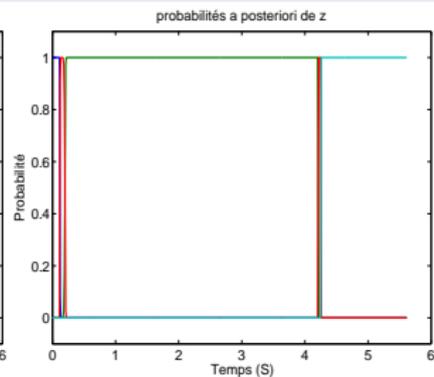
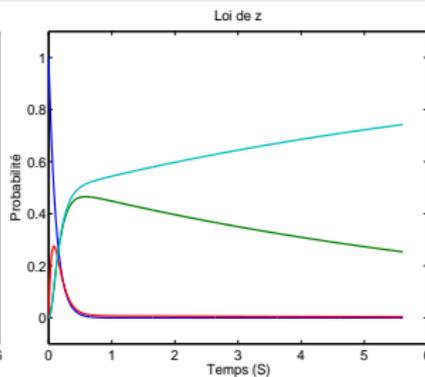
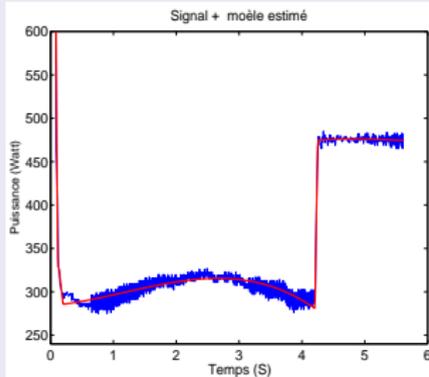
La distribution de \mathbf{z} (ou \mathbf{C}) est :

$$p(\mathbf{z}; \pi, A) = \prod_{k=1}^K p(z_1 = k; \pi)^{z_{1k}} \prod_{i=2}^n \prod_{\ell=1}^K \left(\prod_{k=1}^K p(z_i = k | z_{i-1} = \ell; A)^{z_{ik}} \right)^{z_{(i-1)\ell}} \quad (6)$$

avec :

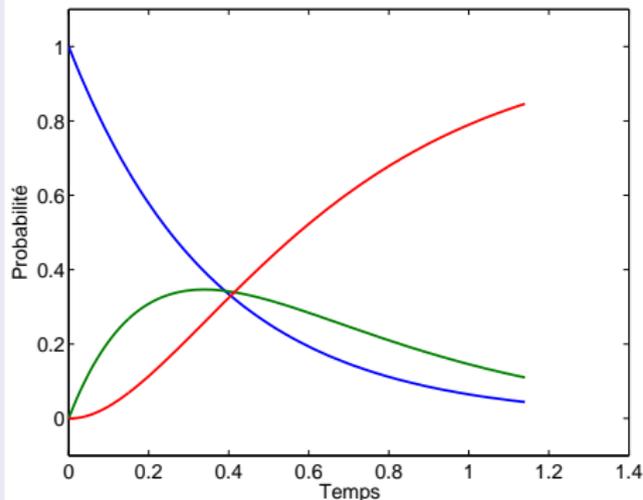
- $z_{ik} = 1$ if $z_i = k$ and 0 otherwise
- $z_{(i-1)\ell} = 1$ if $z_{i-1} = \ell$ and 0 otherwise

Premiers résultats obtenus avec le HMM

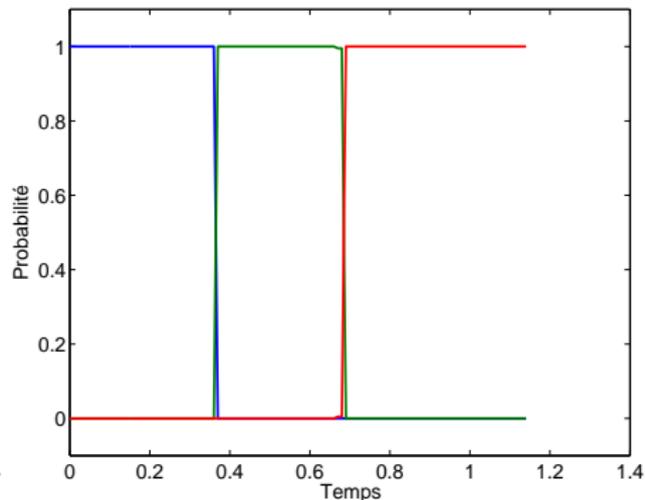


Résultats obtenus avec le HMM pour le suivi de point de fonctionnement.

proba des états du système en fonction du temps



probabilités a posteriori des états du système



Conclusion

Avantages du modèle probabiliste de régression à processus caché par rapport une approche de régression classique :

- Autorise les paramètres régression **évoluer aux cours du temps**
→ meilleure modélisation des signaux
- Particularité de l'algorithme de type EM permettant d'estimer les paramètres du modèle :
 - inclut une procédure de régression logistique pour le calcul des proportions du mélange, qui évoluent au cours du temps
 - moyennes des composantes : courbes polynomiales
- résultats encourageants

Perspectives

- 1 Étude comparative avec d'autres approches (HMM ...)
- 2 Choix de critère de comparaison.
- 3 Suivi séquentiel (en ligne) du point de fonctionnement



P. Dempster, N. M. Laird, and D. B. Rubin.

Maximum likelihood from incomplete data via the EM algorithm (with discussion).
Journal of the Royal Statistical Society B , pages 39, PP. 138, 1977.



A. X. Carvalho and M. A. Tanner.

Modeling nonlinearities with mixtures of-experts of time series models.
International Journal of Mathematics and Mathematical Sciences, 2006.



Krishnapuram, B. ; Carin, L. ; Figueiredo, M.A.T. ; Hartemink, A.J.

Sparse multinomial logistic regression : fast algorithms and generalization bounds
IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 27, Issue 6, Page(s) : 957 - 968
June 2005



P. Green.

Iteratively Reweighted Least Squares for Maximum Likelihood Estimation, and some Robust and Resistant Alternatives.
Journal of the Royal Statistical Society B ,46(2), pp149-192, 1984.



Rabiner, L. R.

A tutorial on hidden Markov models and selected applications in speech recognition.
Proceedings of the IEEE, 77 (2), p. 257286, February 1989.



Baum, L.E., Petrie, T., Soules, G. and Weiss, N.

A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains.
Annals of Mathematical Statistics, 41, p. 164171. 1970.



G. Celeux, J. S. Marques and J. Nascimento

Learning Switching Dynamic Models for Objects Tracking.
Rapport de recherche INRIA, Juillet 2003.



K. Chen, L. Xu, H. Chi

Improved Learning Algorithms for Mixture of Experts in Multiclass Classification.
IEEE Transactions on Neural Networks, Volume 12, Issue 9, Page(s) : 1229 - 1252 November 1999.

Merci pour votre attention !