Apprentissage d'un vocabulaire symbolique pour la détection d'objets dans une image

Sébastien Gadat

gadat@cmla.ens-cachan.fr

CMLA, ENS de Cachan

17 décembre 2004

I - Introduction

Problématique classique de la reconnaissance de formes

- On s'intéresse à un phénomène aléatoire

  $\hookrightarrow$ Exemple : observation d'une scène par une caméra
- Signal d'entrée $\longmapsto$ réalisation I d'une variable aléatoire

  $\hookrightarrow$ Image 20×20 en 256 niveaux de gris
- Codage de I: extraction de variables $\delta_1(I), \ldots \delta_f(I)$

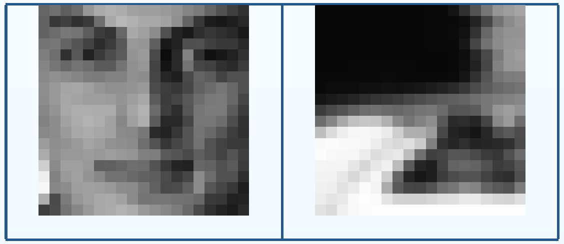
  $\hookrightarrow$ $\delta(I) =$ niveau de gris en un pixel donné
- Interprétation de I

  $\hookrightarrow$Classification d'une image visage / fond

Objectifs:

- Étant donné un signal I, trouver la classe $\mathcal{C}_j(I)$
- Classer l'efficacité des variables $\delta_{i_1}, \ldots \delta_{i_f}$
- Construire un dictionnaire de varaibles utiles pour diverses tâches de reconnaissance

I - Exemples

	Détection visages	Détection de SPAM
I :		"Vahid, à moins de gagner sur tapis vert, pas la peine d'aller jouer dimanche" | "Jouer et gagner facilement au black-jack sur nos tapis verts! Casino on Line"
δ :	Niveau gris	Occurences de mots: "Vahid", "tapis", "!"
But1	$Visage/fond$	$email/Spam$
But2	Pixels informatifs yeux, bouches, nez ...	Mots caractéristiques "Vahid", "!"," Casino" Mots moins importants: "gagner", "tapis"

I - Motivations

Amélioration de performances

- Capacité de stockage
- Vitesse de traitement avec peu de variables
- Certaines variables (non-intéressantes) sont source de bruit

Interprétation de la sélection des variables

- Application en vision:
  - Identification des zones informatives pour une image
- Applications en biologie (génétique) pour la sélection de variables:
  - Les variables: localisées sur certains gènes
  - En général, deux classes: Sains/Malades.
  - Importance de l'identification des gènes pour élaboration médicaments
  - Sélection des gènes *via* extraction variables

I - Objectifs pour l'extraction de variables

$\mathcal{F}_0$ est l'ensemble des variables (tests, questions, ...) initiales δ applicables sur le signal

1. Trouver sous-ensemble de $\mathcal{F}_0$ restreint

   $\hookrightarrow$ Existe-t-il un sous-ensemble de $\mathcal{F}_0$ suffisant pour décrire les données?

2. Enrichir $\mathcal{F}_0$

   $\hookrightarrow$ Peut-on former de nouvelles variables pour le problème de classification?

3. Formalisation
   - recherche d'un sous-espace de $\mathcal{F}_0$
   - proposer une sélection de variables

I - Cadre de travail

1. Cadre classification supervisé, *i.e.* on dispose
   - Échantillons exemples (Training Set) pour chaque classe
     $\hookrightarrow$ Exemple: images classées dans $\{0, \ldots 9\}$ en grand nombre
   - Échantillons à classer (Test Set pour évaluer)

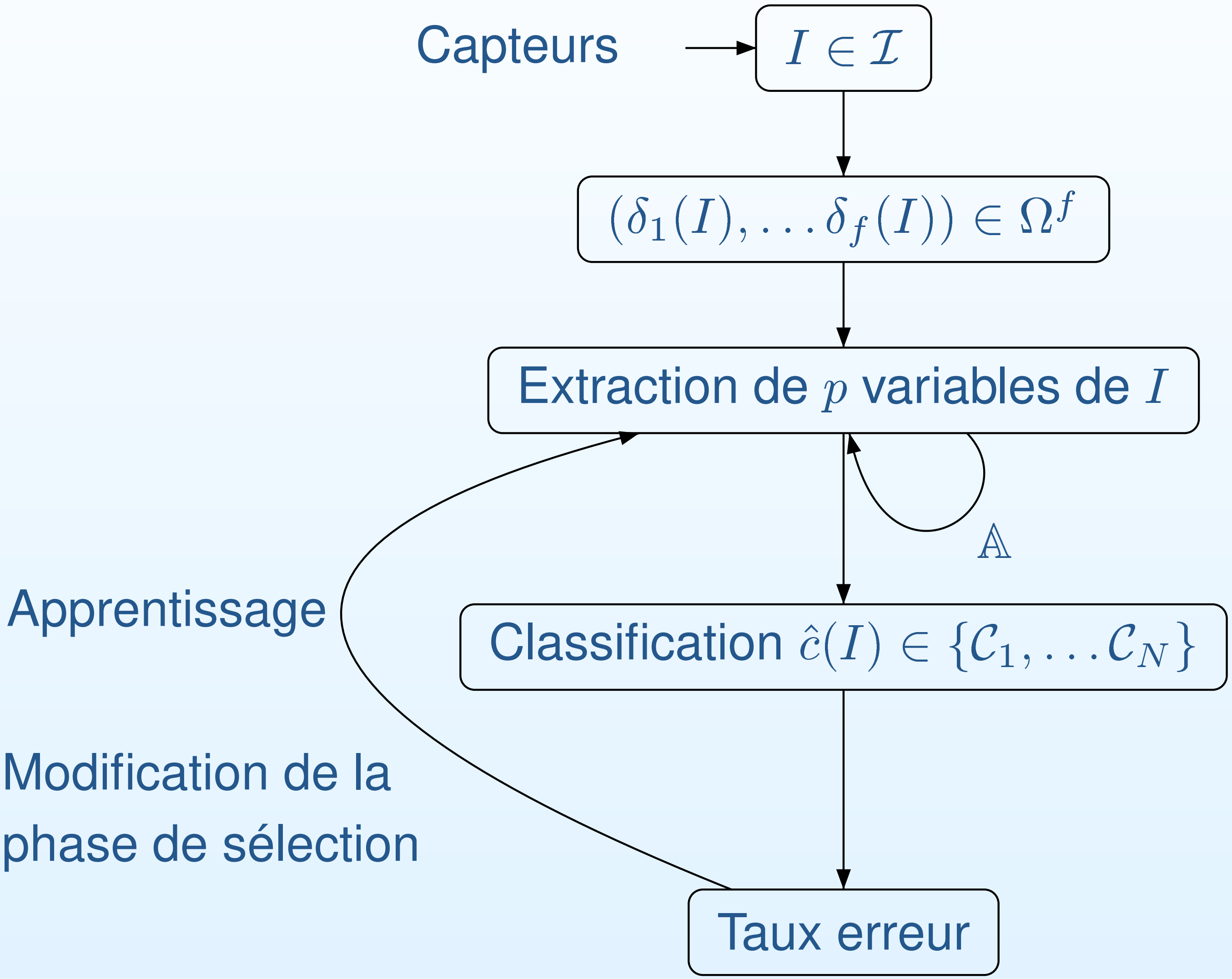
2. Variables δ très rapides à calculer sur un signal I
   - Images: Détecteurs de bords binaires/ternaires
   - Valeurs réelles pour autres types signaux (emails, . . .)
   - δ à valeurs dans Ω

3. Outils de base: Algorithmes de classification notés $\mathbb{A}$
   - Signal $I \longmapsto$ vecteur $(\delta_1(I), \ldots \delta_f(I)) \in \Omega^f$
   - $\mathbb{A}$ s'applique sur sous-signaux de Ω^p, $p << f$
   - $\mathbb{A}$ difficilement applicable aux signaux entiers de Ω^f car:
     - Temps calcul / Mauvaises variables: bruitent données

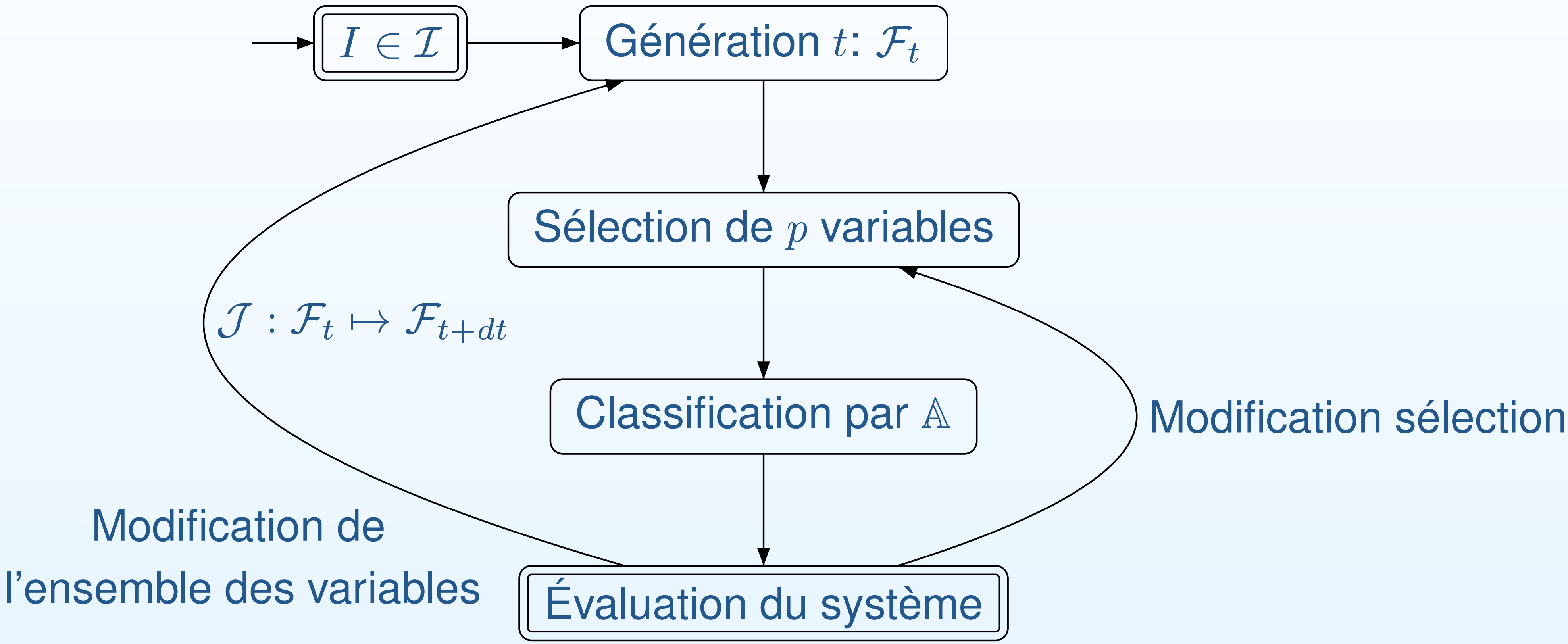
I - Notations

- Signal mesuré I
- Variables explicatives (tests) sur I notées δ
- Ω: réalisation des tests
- Variables initiales notées $\mathcal{F}_0$
- Ensemble de variables évoluant avec t noté $\mathcal{F}_t$
- $\mathbb{P}_t$ probabilité de support $\mathcal{F}_t$ à l'instant t
- Parcourt de l'ensemble des variables et probabilités: $\mathcal{F}_t$ et $\mathbb{P}_t$ évoluent
- Améliorations pour le traitement du signal I tout en gardant une quantité de variables restreint

I - Schéma de l'apprentissage de la sélection de variables



I - Schéma de l'apprentissage de l'espace des tests $\mathcal{F}_t$



I - Algorithmes $\mathbb{A}$

1. Paramètres d'entrées: sous-ensemble de variables explicatives et signaux étiquettés (*Training Set* $*$ et $\times$)
2. Signal I à classer: $\bigstar$
3. Utilisation des deux algorithmes:

4. Sortie: une classe $\times$ ou $*$

II - Extraction de variables par minimisation d'énergie

On suppose $\mathcal{F}_t$ fixé $(= \mathcal{F}_0)$

1. Loi $\mathbb{P}$ sur $\mathcal{F}$

   $\hookrightarrow$ tirage avec remise de variables dans $\mathcal{F}_0$

2. Mesure d'erreur $g(\omega)$ pour $\omega \in \mathcal{F}_0^p$

   $\hookrightarrow$ $g(\omega)$ mesure l'erreur connaissant les variables de ω

   - $N = 2$, deux classes

     $$g(\omega) = P\left[\mathbb{A}(\omega(X)) \neq \mathcal{C}(X) \quad | \quad X \in Training\, Set\right]$$

   - $N \geq 3$, multi-classes, g_i: erreur relative à $(\mathcal{C}_i | \cup_{j \neq i} \mathcal{C}_j)$

     $$g(\omega) = \sum g_i(\omega)$$

3. Énergie $\mathcal{E}_1(\mathbb{P})$

   $\hookrightarrow$ à $\mathbb{P}$ est associée une mesure d'efficacité moyenne:

   $$\mathcal{E}_1(\mathbb{P}) = \sum_{\omega \in \mathcal{F}_0{}^p} g(\omega)\mathbb{P}(\omega) = \mathbb{E}_{\mathbb{P}}\left[g(\omega)\right]$$

II - Modélisation, rappel vers la loi uniforme

1. Propriétés de $\mathbb{P}$ optimale recherchée
   - Quantifier utilité des variables $\delta \in \mathcal{F}_0$ par $\mathbb{P}(\delta)$
     $\hookrightarrow$ Minimisation d'une énergie basée sur $\mathcal{E}_1$
   - Tirage de p tests par $\mathbb{P}$
     $\hookrightarrow$ collaboration de plusieurs variables pour un signal
   - Support de $\mathbb{P}$ non dégénéré à un petit ensemble de variables
     $\hookrightarrow$ Utilisation d'une procédure de vote efficace pour détection
2. Énergie de régularisation $\mathcal{E}_2$: rappel vers loi uniforme sur $\mathcal{F}_0$

$$\mathcal{E}_2(\mathbb{P}) = \left\|\mathcal{U}_{\mathcal{F}_0} - \mathbb{P}\right\|_2^{\,2}$$

   - Rappel vers $\mathcal{U}_{\mathcal{F}_0}$ pour régularisation de l'algorithme
   - minima locaux non-dégénérés
3. Énergie globale:

$$\mathcal{E}(\mathbb{P}) = \mathcal{E}_1(\mathbb{P}) + \lambda\mathcal{E}_2(\mathbb{P}) \qquad \lambda > 0$$

II - Méthode d'optimisation

- Énumération des $g(\omega)$ incalculables, optimisation par descente de gradient sur $\mathbb{P}$
- Descente de gradient classique

  $\hookrightarrow$ Gradient métrique euclidienne

  $\hookrightarrow$ Respect des contraintes type "simplexe"

$$\mathcal{S}_{\mathcal{F}_0} : \begin{cases} \mathbb{P}(\delta_1) + \cdots + \mathbb{P}(\delta_f) = 1 \\ \mathbb{P}(\delta_i) \geq 0 \end{cases}$$

- Descente de gradient variables exponentielles $y = (y_1, \ldots y_f)$

$$\mathbb{P}(\delta_i) = e^{y_i} \qquad \delta_i \in \mathcal{F}_0$$

  $\hookrightarrow$ Facilité respect contraintes

  $\hookrightarrow$ Évolution plus lente (frontière $\partial\mathcal{S}_{\mathcal{F}_0}$ à distance infinie)

  $\hookrightarrow$ Descente de gradient métrique de Kullback-Leibler

II - Approximation stochastique

- Schémas de descente

$$(\mathbf{E}) \begin{cases} \mathbb{P}_{n+1} = \mathbb{P}_n - \epsilon \nabla_{\mathcal{S}_{\mathcal{F}_0}} \mathcal{E}(\mathbb{P}_n) \\ y_{n+1} = y_n - \epsilon \nabla_{\mathcal{S}_{\mathcal{F}_0}} \mathcal{E}(y_n) \end{cases}$$

$\hookrightarrow$ On approche la solution de

$$\frac{\partial \mathbb{P}_t}{\partial t} = -\nabla_{\mathcal{S}_{\mathcal{F}_0}} \mathcal{E}(\mathbb{P}_t)$$

- Implémentation de la descente approchée classique impossible:

$$\nabla \mathcal{E}(\mathbb{P})(\delta) = \sum_{\omega} g(\omega) C(\omega, \delta) \frac{\mathbb{P}(\omega)}{\mathbb{P}(\delta)}$$

où $C(\omega, \delta)$ est le nombre d'apparitions de δ dans le p-uplet ω
$\hookrightarrow$ On effectue une approximation stochastique de $(\mathbf{E})$

$$\mathbb{P}_{n+1} = \mathbb{P}_n + \gamma_n d_n \qquad \text{ou} \qquad \mathbb{P}_{n+1} = \mathbb{P}_n \frac{e^{\gamma_n d_n}}{K_n}$$

avec $\mathbb{P}_0 = \mathcal{U}_{\mathcal{F}_0}$ loi uniforme sur $\mathcal{F}_0$

$\hookrightarrow$ Trouver réels $(\gamma_n)_{n \in \mathbb{N}}$ et vecteurs $(d_n)_{n \in \mathbb{N}}$ approchant $(\mathbf{E})$

II - Approximation de Robbins-Monro

- Conditions théoriques sur $(\gamma_n)_{n\in\mathbb{N}}$:

$$\begin{cases} \sum \gamma_n = +\infty \\ \sum \gamma_n^2 < +\infty \end{cases}$$

$\hookrightarrow$ Permet de donner une discrétisation du temps réel en terme d'itérations n:

$$\tau_k = \sum_{i=1}^{k} \gamma_i \in \mathbb{R}^+$$

- Condition sur d_n:

$$\mathbb{E}_{\mathbb{P}_n}[d_n] = -\nabla_{\mathcal{S}_{\mathcal{F}_0}} \mathcal{E}(\mathbb{P}_n)$$

$\hookrightarrow$ L'effet moyen de d_n est identique à la descente de gradient exacte

II - Convergence de l'approximation

- Les suites $(\mathbb{P}_n)_{n\in\mathbb{N}}$ et $(y_n)_{n\in\mathbb{N}}$ convergent vers des solutions de

$$\begin{cases} \frac{\partial \mathbb{P}}{\partial t} = -\nabla_{\mathcal{S}_{\mathcal{F}_0}} \mathcal{E}(\mathbb{P}_t) \\ \frac{\partial y}{\partial t} = -\nabla_{\mathcal{S}_{\mathcal{F}_0}} \mathcal{E}(y_t) \end{cases}$$

- Difficulté pour restreindre $(\mathbb{P}_n)_{n\in\mathbb{N}}$ à $\mathcal{S}_{\mathcal{F}_0}$ par

$$\mathbb{P}_{n+1} = \mathbb{P}_n + \gamma_n d_n$$

  Critère déterministe:

$$\lambda > 2pf^2 \sup g(\omega)$$

- Variables exponentielles: lenteur d'évolution mais facilité respect contraintes
- Variables en $\mathbb{P}$: rapidité mais difficulté du respect des contraintes
- Autres termes de rappel: $\mathcal{E}_2(\mathbb{P}) = -H(\mathbb{P})$

  $\hookrightarrow$ Attraction vers $\mathcal{U}$ trop forte: difficulté pour régler λ

II - Expériences - Données synthétiques

- 100 variables synthétiques ternaires, 3 classes $\mathcal{C}_i$:
  $\mathcal{I} = \{-1;0;1\}^{100}$
- pour $\mathcal{G} \subset \mathcal{F}_0$: $\mu(\delta, \mathcal{G}) \sim \mathcal{U}_{\{-1;0;1\}} \chi_{\delta \notin \mathcal{G}} + \delta_{\{1\}} \chi_{\delta \in \mathcal{G}}$
- on génère les données de $\mathcal{C}_i$ en utilisant les distributions

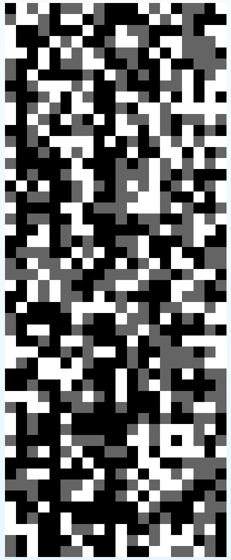
$$\mu_i(I) = \frac{9}{10} \frac{\mu(I,\mathcal{G}_i^1)+\mu(I,\mathcal{G}_i^2)+\mu(I,\mathcal{G}_i^3)}{3} + (1 - \frac{9}{10})\mathcal{U}_{\{-1;0;1\}}$$

- Tests utilisés pour générer les classes $\mathcal{C}_i$:

$$\begin{array}{lll} \mathcal{G}_1^1 = \{\delta_1, \delta_3, \delta_5, \delta_7\} & \mathcal{G}_1^2 = \{\delta_1, \delta_5\} & \mathcal{G}_1^3 = \{\delta_3, \delta_7\} \\ \mathcal{G}_2^1 = \{\delta_2, \delta_4, \delta_6, \delta_8\} & \mathcal{G}_2^2 = \{\delta_2, \delta_4\} & \mathcal{G}_2^3 = \{\delta_6, \delta_8\} \\ \mathcal{G}_3^1 = \{\delta_1, \delta_4, \delta_8, \delta_9\} & \mathcal{G}_3^2 = \{\delta_1, \delta_8\} & \mathcal{G}_3^3 = \{\delta_4, \delta_9\} \end{array}$$

II - Expériences - Données synthétiques

- Échantillon

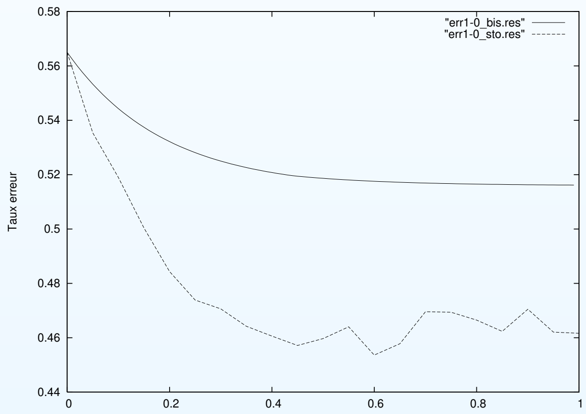
$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}_3$
		

- $\mathbb{A}$: 4 plus proches voisins avec la distance

$$d(I_1, I_2) = \sum_{\delta} \chi_{\delta(I_1) \neq \delta(I_2)}$$

II - Expériences - Données synthétiques

- Comparaison erreur moyenne $\mathcal{E}_1$ descente exacte, stochastique, variables en $\mathbb{P}$:



- Probabilité après apprentissage

δ_1	δ_2	δ_3	δ_4	δ_5	δ_6	δ_7	δ_8	δ_9	δ_{10}	δ_{11}	δ_{12}	δ_{13}	δ_{14}

- Variables sélectionnées: dans intersection de $\mathcal{G}_i^j$

II - Expériences - Détection de visages

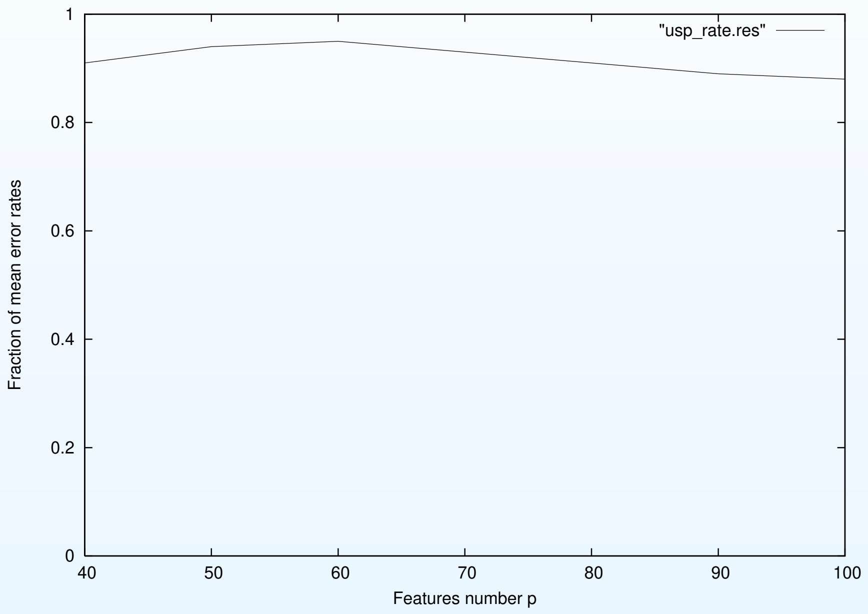
- Imagettes 19×19, niveaux de gris $\{0 \ldots 255\}$, centrées
- Variables élémentaires de $\mathcal{F}_0$

  $\hookrightarrow$ $\mathcal{F}_0$: détecteurs de bords variables binaires, $|\mathcal{F}_0| \simeq 2000$

  $\hookrightarrow$ $\mathcal{F}_0$: ensemble des niveaux de gris
- $\mathbb{A}$: algorithme de Support Vector Machine, $p = 20 - 100$
- Paramétrisation de $\mathbb{P}$ en variables exponentielles
- Échantillons

II - Expériences - Détection de visages

- Rapport $\frac{\mathcal{E}_1(\mathbb{P}_\infty)}{\mathcal{E}_1(\mathcal{U}_{\mathcal{F}_0})}$ en fonction de p

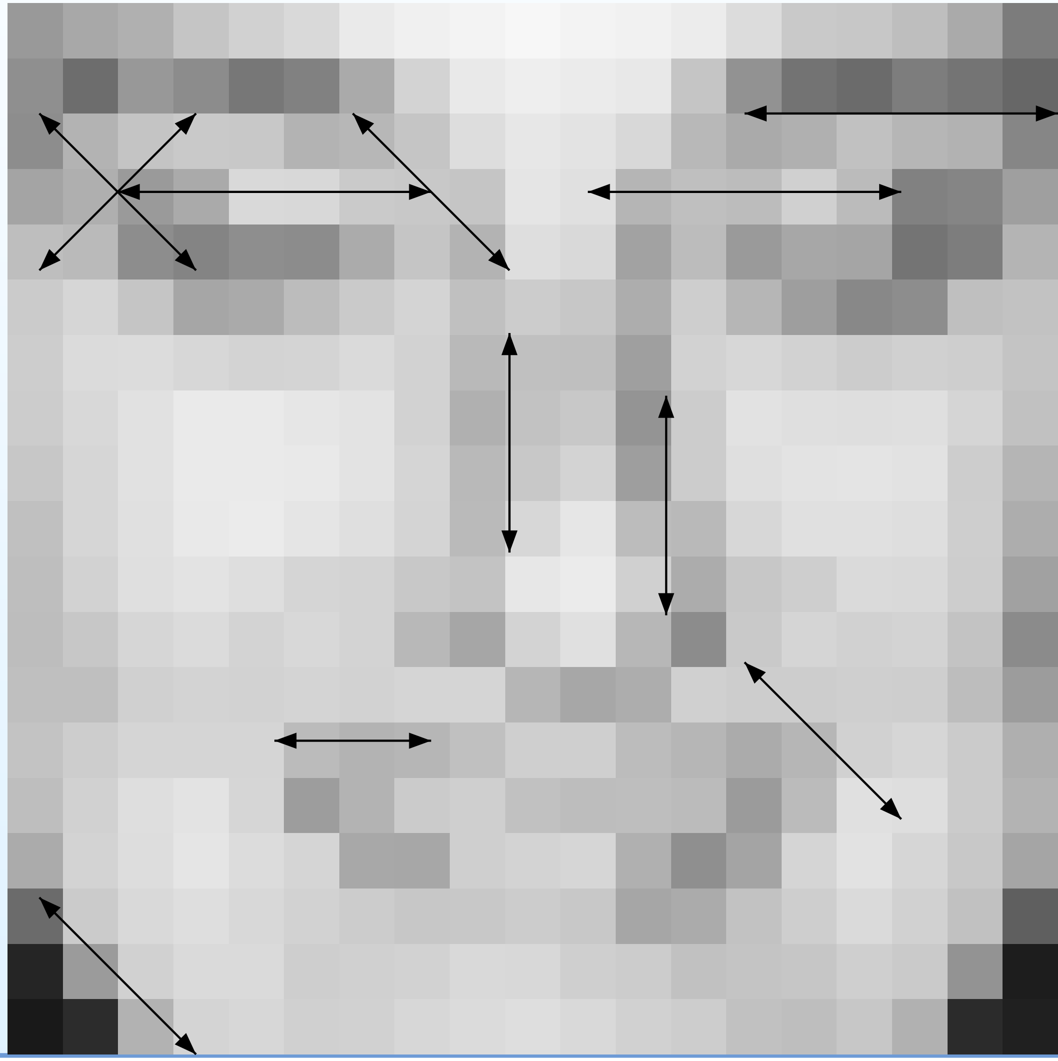


- Performance sur le Test-Set par vote détecteurs:

Nombre variables	$\mathcal{U}_{\mathcal{F}_0}$	$\mathbb{P}_\infty$
$p = 20$	13.6%	7.5%
$p = 100$	3.5%	2.3%

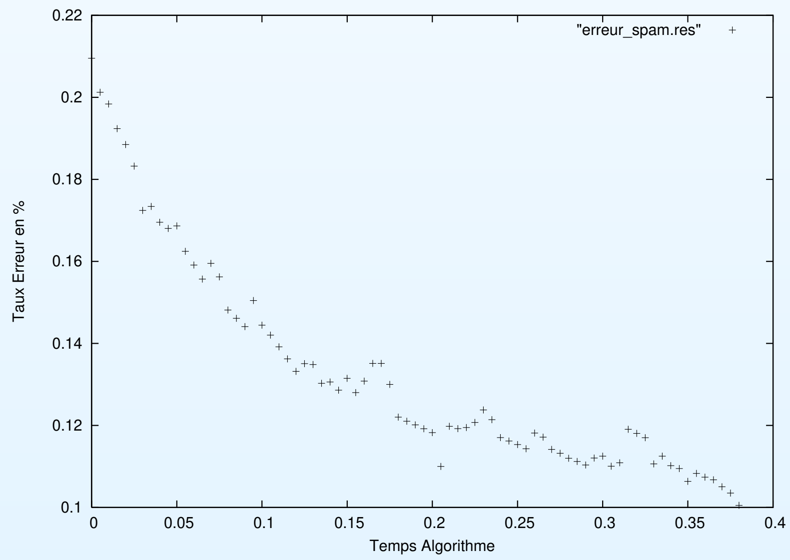
II - Expériences - Détection de visages

- Amélioration des performances moyenne / méthode naïve
- Représentation des détecteurs de bords privilégiés:



II - Expériences - Classification de SPAM

- I: message électronique, $\mathcal{F}_0$: pourcentage d'occurrences de mots dans un message (liste de 54 mots prédéfinie)
- $\mathbb{A}$: 4 plus proches voisins
- $p = 10 - 15$ mots tirés selon $\mathbb{P}$
- Bonne décroissance de l'erreur moyenne $\mathcal{E}_1$ au cours du temps:



II - Expériences - Classification de SPAM

- Après vote de détecteurs: taux d'erreur de 7.5%
- Sélection de mots par $\mathbb{P}_\infty$:

Mots - SPAM	Fréquence	Mots -NON SPAM	Fréquence
report	8.8%	cs	5.4%
business	8.7%	857	4.6%
[	6%	415	4.4%
remove	5.9%	project	4.3%
receive	5.6%	table	4.2%
internet	4.4%	conference	4.2%
free	4.1%	lab	3.9%
people	3.7%	labs	3.2%
000	3.6%	edu	2.8%
direct	2.3%	650	2.7%
!	1.2%	85	2.5%
$	1%	george	1.6%

III - Extraction d'un espace de variables

Modèle précédent:

$$\mathcal{E}(\mathbb{P}) = \underbrace{\mathbb{E}_{\mathbb{P}}\left[g(\omega)\right]}_{\mathcal{E}_1(\mathbb{P})} + \lambda \underbrace{\left\|\mathcal{U}_{\mathcal{F}_0} - \mathbb{P}\right\|_2^2}_{\mathcal{E}_2(\mathbb{P})}$$

- Peut-on faire évoluer l'ensemble de variables $\mathcal{F}$ plutôt que seulement un poids $\mathbb{P}$ sur $\mathcal{F}$?

  $\hookrightarrow$ Garder un système de notations d'efficacité *via* $\mathbb{P}$

  $\hookrightarrow$ Permettre suppression de variables mal notées par $\mathbb{P}$ de $\mathcal{F}$

  $\hookrightarrow$ Permettre ajoût de variables composées, bien notées par $\mathbb{P}$

- Suppression du rappel vers $\mathcal{U}_{\mathcal{F}}$, "aveugle" à l'apprentissage sur $\mathcal{E}_1$ et d'autant plus fort que $\mathbb{P}$ est éloignée de $\mathcal{U}_{\mathcal{F}}$

  $\hookrightarrow$ Trouver un substitut à $\mathcal{E}_2$?

III - Équation d'évolution

$\mathcal{F}_t$ fixé, suppression de $\mathcal{E}_2$, et définition "naturelle" de la stabilité de $\mathcal{S}_{\mathcal{F}}$

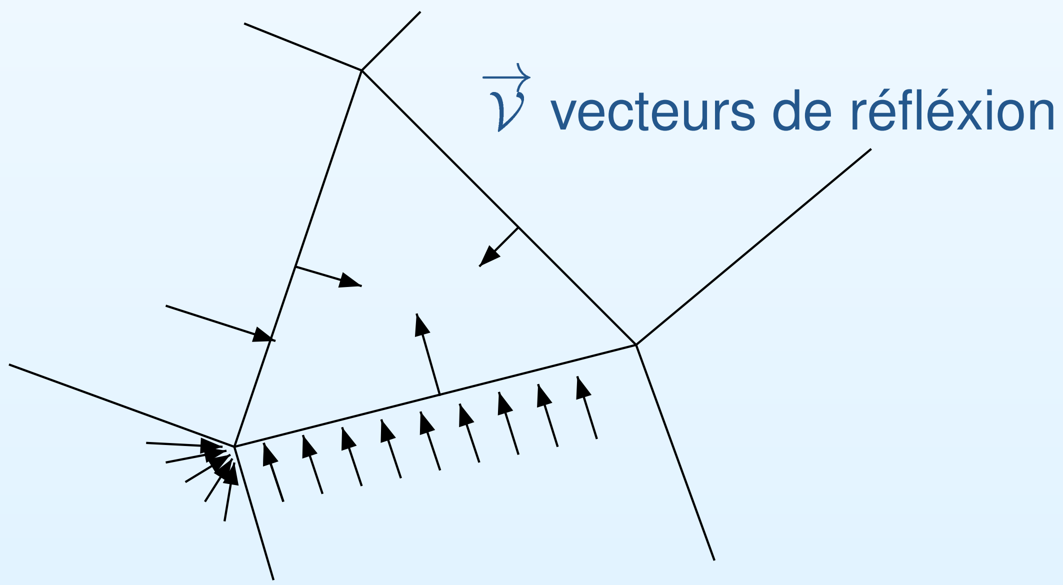
- Utilisation d'une équation différentielle comparable à

$$d\mathbb{P}_t = -\nabla_{\mathcal{S}_{\mathcal{F}}}\mathcal{E}_1(\mathbb{P}_t)dt - \nabla_{\mathcal{S}_{\mathcal{F}}}\mathcal{E}_2(\mathbb{P}_t)dt$$

- Mouvement brownien sur $\mathcal{S}_{\mathcal{F}_t}$:

$$d\mathbb{P}_t = -\nabla_{\mathcal{S}_{\mathcal{F}}}\mathcal{E}_1(\mathbb{P}_t)dt + \sqrt{\sigma}dW_t$$

- Restreindre $(\mathbb{P}_t)$ à $\mathcal{S}_{\mathcal{F}}$: diffusion réfléchie sur le simplexe:



III - Restriction de $\mathbb{P}_t$ à $\mathcal{S}_{\mathcal{F}}$

Utilisation des directions de réflexion

- A chaque instant où $\mathbb{P}_t$ atteint $\partial\mathcal{S}_{\mathcal{F}}$: direction de réflexion

$$d\mathbb{P}_t = -\nabla_{\mathcal{S}_{\mathcal{F}}}\mathcal{E}_1(\mathbb{P}_t)dt + \sigma(\mathbb{P}_t)dW_t + dZ_t \qquad \text{(E')}$$

- Z_t n'est modifié que lorsque $\mathbb{P}_t \in \mathcal{S}_{\mathcal{F}}$

$$|Z_t| = \int_0^t \chi_{\mathbb{P}_s \in \partial\mathcal{S}_{\mathcal{F}}} d|Z_s|$$

$$Z_t = \int_0^t \gamma(s) d|Z|_s \qquad \text{où} \qquad \gamma(s) \in \overrightarrow{\mathcal{V}}$$

Existence et unicité de diffusions réfléchies Sous les conditions

1. $|\sigma(x) - \sigma(y)| + |\nabla\mathcal{E}_1(x) - \nabla\mathcal{E}_1(y) \leq k_1|x-y|$
2. $|\nabla\mathcal{E}_1(x)| \leq k_2(1+|x|)$
3. $|\sigma(x)| \leq k_3$

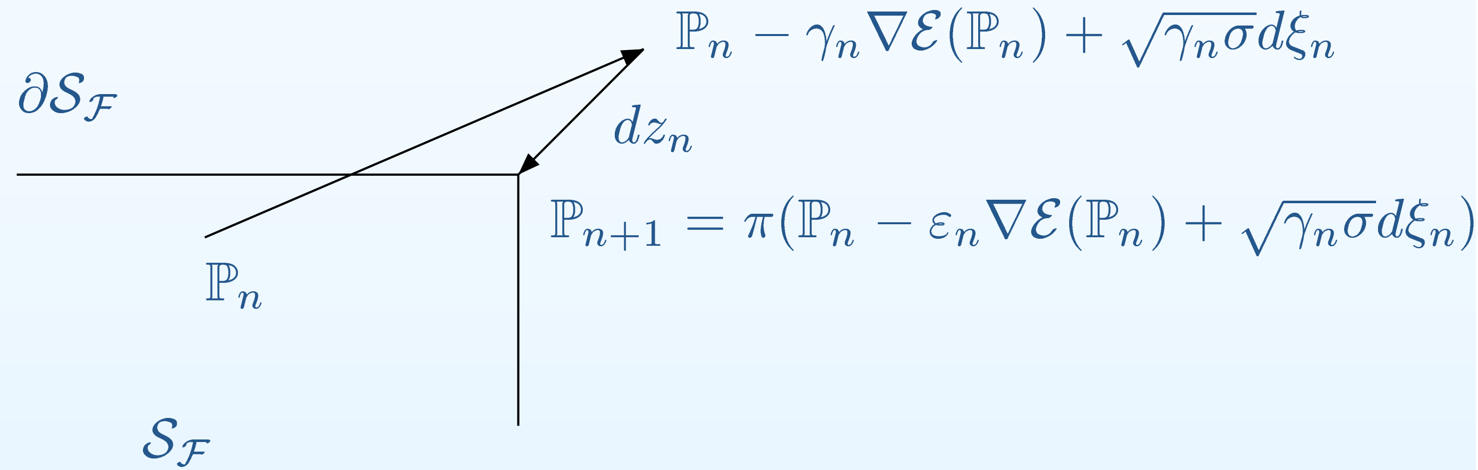
Il existe un unique processus vérifiant (E')

III - Simulation de (E’)

- Définition de $\mathbb{P}_n$:

  $$\mathbb{P}_{n+1} = \mathbb{P}_n - \gamma_n d_n + \sqrt{\gamma_n \sigma} d\xi_n + dz_n = \pi_{\mathcal{S}_\mathcal{F}} \left(\mathbb{P}_n - \gamma_n d_n + \sqrt{\gamma_n \sigma} d\xi_n \right)$$

  où $\mathbb{E}_{\mathbb{P}_n}[d_n] = \nabla_{\mathcal{S}_\mathcal{F}} \mathcal{E}_1(\mathbb{P}_n)$, $d\xi_n$ gaussienne standard, π projection sur $\mathcal{S}_\mathcal{F}$

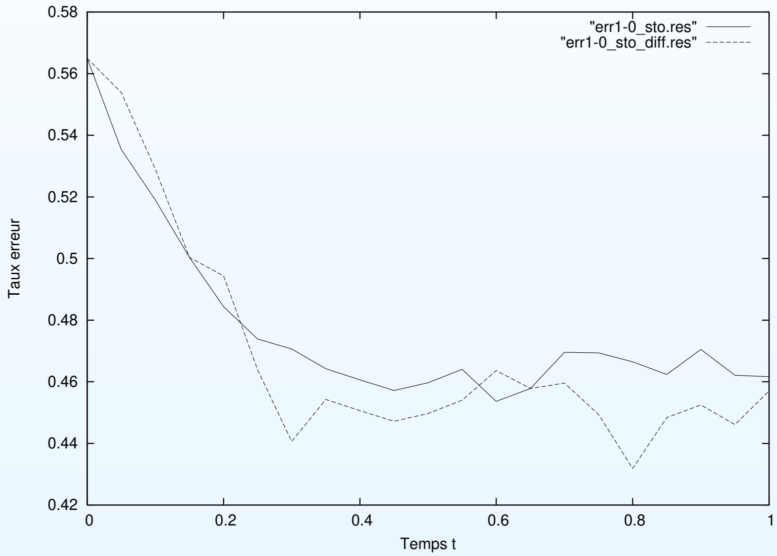


- convergence de $\mathbb{P}_n$ vers diffusion réfléchie si

  $$\sum \gamma_n^{1+\alpha} < \infty$$

III - Expériences Diffusion réfléchie

- Données synthétiques comparable au gradient approchée



- Détection visages Évolution plus lente, taux erreurs comparables

$\lvert\omega\rvert$	Descente de gradient	Diffusion réfléchie
20	7.5%	8%
100	2.3%	2.7%

III - Arbres binaires de variables, opérations

Variables binaires, faire évoluer $\mathcal{F}_t$?

- $\mathcal{F}^\star$ ensemble des arbres binaires constructibles par:

$$\delta \in \mathcal{F}_0 \Longrightarrow \delta \in \mathcal{F}^\star$$

$$(\delta.g, \delta.d) \in \mathcal{F}^{\star 2} \Longrightarrow \quad (\delta.g) \times (\delta.d) \in \mathcal{F}^\star$$

$$\delta.g \qquad \delta.d$$

- Règles de parcours de $\mathcal{F}^\star$ par $\mathcal{F}_t$:
  1. $T_g(\mathcal{A}_1; \mathcal{A}_2)$ Greffe de $\mathcal{A}_1$ et $\mathcal{A}_2$, suppression possible de fils
  2. T_c Coupe d'un arbre en deux sous-arbres
  3. T_i Renaissance d'un arbre de $\mathcal{F}_0$
- Avec T_g, T_c et T_i, $(\mathcal{F}_t)_{t \geq 0}$ peut explorer $\mathcal{F}^\star$, sa dynamique est faiblement réversible.
- X_t indicateur des arbres dans $\mathcal{F}_t$: $X_t(\mathcal{A}) = \chi_{\mathcal{A} \in \mathcal{F}_t}$

III - Énergie du système

- On définit pour le système $(\mathcal{F}_t, \mathbb{P}_t)$, initialisé en $(\mathcal{F}_0, \mathbb{P}_0)$:

$$\tilde{\mathcal{E}}(\mathcal{F}_t, \mathbb{P}_t) = e^{-(\mathcal{E}_1(\mathbb{P}_t)+\mathcal{E}_{Lg}(\mathcal{F}_t)+\mathcal{E}_\rho(\mathcal{F}_t))}$$

  - $\mathcal{E}_1$ énergie d'erreur moyenne
  - $\mathcal{E}_{Lg}$ terme de complexité

$$\mathcal{E}_{Lg}(\mathcal{F}_t) = \sum_{\mathcal{A}\in\mathcal{F}_t} |\mathcal{A}|$$

  - $\mathcal{E}_\rho$ terme issu de la théorie de l'information:

$$\mathcal{E}_\rho(\mathcal{F}_t) = -\sum_{\mathcal{A}\in\mathcal{F}_t} I(\mathcal{A}.g, \mathcal{A}.d)$$

- Énergie $\tilde{\mathcal{E}}$: conditionne les sauts dans l'espace des variables $\mathcal{F}^\star$
- Acceptation de sauts selon la règle de Métropolis Hastings

$$Q\begin{pmatrix} \mathcal{F}_{t_s} & ; & \mathcal{F}_{t_s+dt} \\ \mathbb{P}_{t_s} & & \mathbb{P}_{t_s+dt} \end{pmatrix} = \min\left\{1; \frac{P(\mathcal{F}_{t_s+dt}|\mathcal{F}_{t_s})}{P(\mathcal{F}_{t_s}|\mathcal{F}_{t_s+dt})} e^{-\Delta(\tilde{\mathcal{E}})}\right\}$$

III - Équation différentielle du système couplé

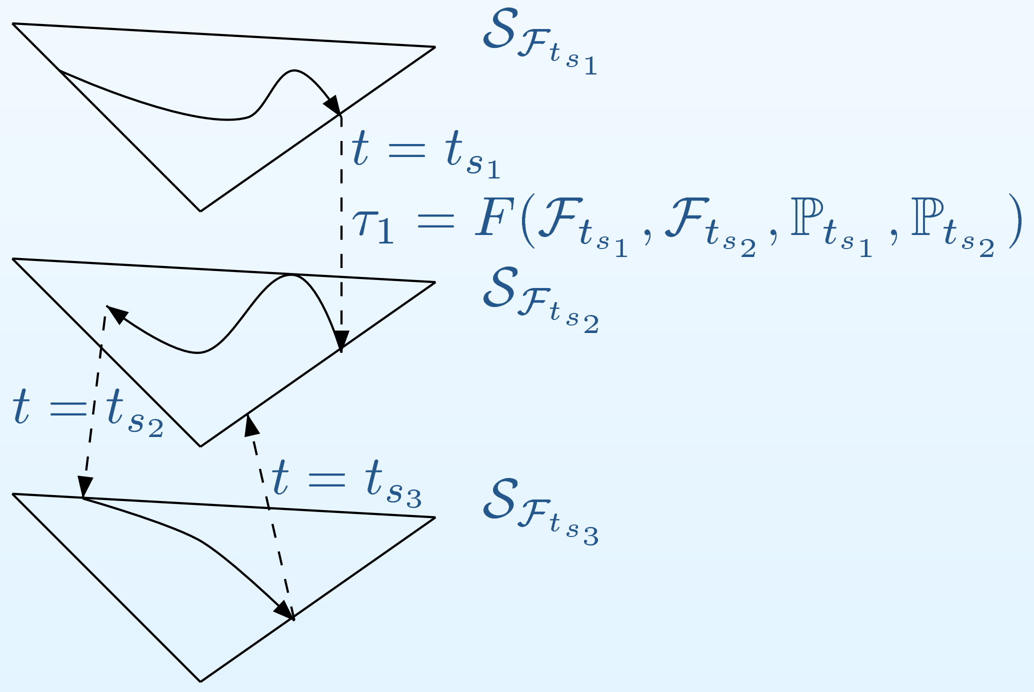
- Processus $(\mathbb{P}_t, X_t)$ défini par:

$$(E'') \quad d\begin{pmatrix} \mathbb{P}_t \\ X_t \end{pmatrix} = -\begin{pmatrix} \nabla\mathcal{E}_1(\mathbb{P}_t) \\ 0 \end{pmatrix} dt + \begin{pmatrix} \sigma(X_t, \mathbb{P}_t) \\ 0 \end{pmatrix} dW_t + \int Q\begin{pmatrix} \mathbb{P}_t & ; & \mathbb{P} \\ X_t & & X \end{pmatrix} N\left(d\begin{pmatrix} \mathbb{P} \\ X \end{pmatrix}; dt\right) + \begin{pmatrix} dZ_t \\ 0 \end{pmatrix}$$

- Q probabilité de transition précédente
- N distribution des sauts selon une loi de Poisson en temps, et uniforme en espace
- dZ_t direction de réflexion de rappel
- X_t et $\mathcal{F}_t$, ne sont modifiés qu'aux instants de sauts donnés par N

III - Existence du processus

- il existe un unique processus continu à droite vérifiant (E'')
- instants de sauts $(t_{s_i})_{i\in\mathbb{N}}$ distribués selon $\mathcal{P}(\lambda)$
- pour $t \in [t_{s_i}; t_{s_{i+1}}[$ diffusion réfléchie sur $\mathcal{S}_{\mathcal{F}_{t_{s_i}}}$
- instants t_{s_i}: transition $\mathcal{F}_{t_{s_i}} \mapsto \mathcal{F}_{t_{s_i}+dt}$ conditionnée par $\Delta\tilde{\mathcal{E}}$



III - Dynamique du processus

- Le processus défini par (E'') est récurrent positif: le temps d'atteinte de tout compact S de mesure strictement positive vérifie:

$$\forall(\mathcal{F};\mathbb{P}) \in (\mathcal{F}^\star;\mathcal{S}_{\mathcal{F}^\star}) \qquad \mathbb{E}_{(\mathcal{F};\mathbb{P})}\tau_S < \infty$$

- L'unique mesure stationnaire du processus est le champ de Gibbs associé à $\tilde{\mathcal{E}}$

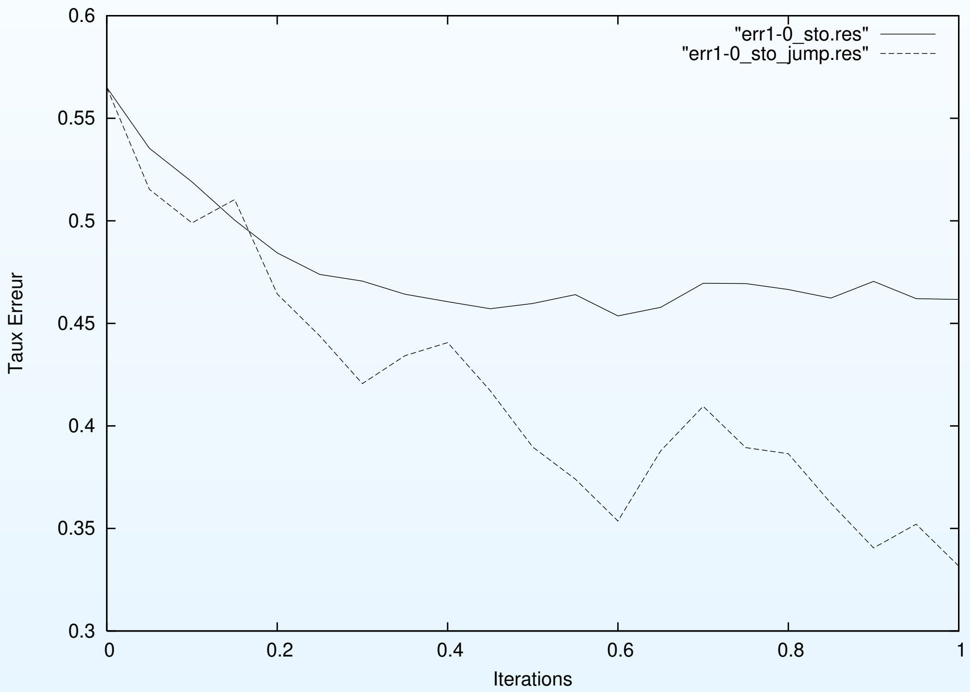
$$\mu(\mathbb{P},\mathcal{F}) = \frac{e^{-\tilde{\mathcal{E}}(\mathbb{P},\mathcal{F})}}{Z}$$

III - Approximation de (E'')

- Gradient de $\mathcal{E}_1$ numériquement incalculable
- Approximation identique au gradient stochastique du II
- Convergence au sens faible du processus interpolé vers (E”), et vers l’unique mesure stationnaire de (E”)

III - Expériences - Données synthétiques

- Comparaison Diffusion avec et sans processus de sauts



- Évolution plus irrégulière, taux erreurs inférieurs en utilisant le processus de sauts

III - Expériences - Données synthétiques

- Nombre d'arbres réduits
- Mesure d'occupation des arbres

$$\mu_t(\mathcal{A}) = \int_0^t \chi_{\mathcal{A} \in \mathcal{F}_s} \, ds$$

- Par ordre décroissant de présence

$$\{\delta_2 ; \delta_4\} \quad \{\delta_1 ; \delta_5\} \quad \{\delta_4 ; \delta_9\} \quad \{\delta_1 ; \delta_8\} \quad \{\delta_6 ; \delta_8\} \quad \{\delta_3 ; \delta_7\} \quad \delta_1 \quad \delta_4 \quad \delta_8$$

$$\{\delta_1 ; \delta_5 ; \delta_3\} \quad \{\delta_2 ; \delta_4 ; \delta_8\} \quad \{\delta_4 ; \delta_9 ; \delta_8\}$$

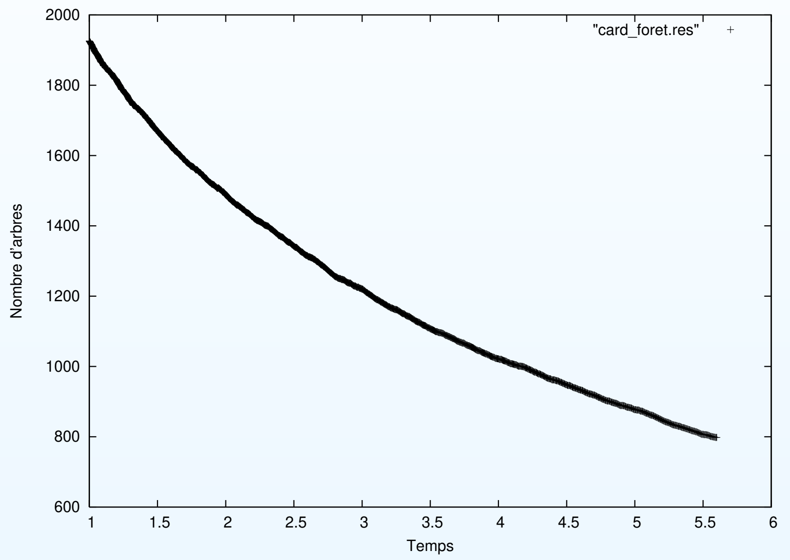
- Sources initiales $\mathcal{G}_i^j$ quasiment reconstituées

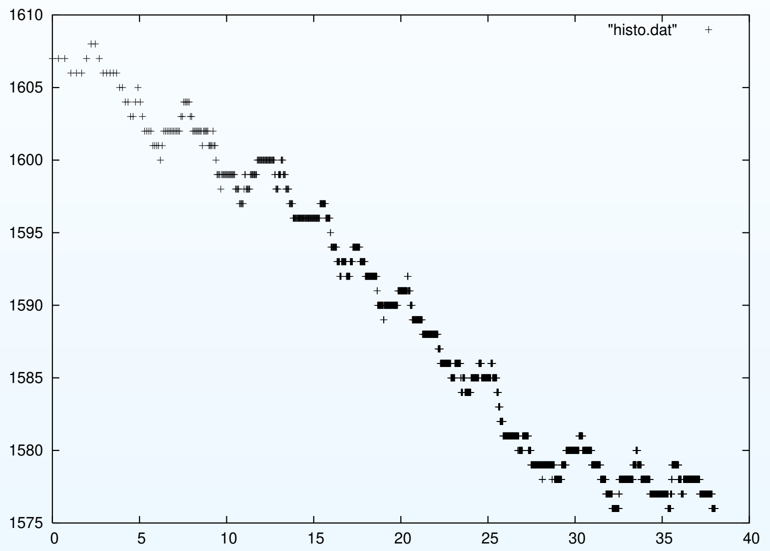
$$\mathcal{G}_1^1 = \{\delta_1, \delta_3, \delta_5, \delta_7\} \quad \mathcal{G}_1^2 = \{\delta_1, \delta_5\} \quad \mathcal{G}_1^3 = \{\delta_3, \delta_7\}$$

$$\mathcal{G}_2^1 = \{\delta_2, \delta_4, \delta_6, \delta_8\} \quad \mathcal{G}_2^2 = \{\delta_2, \delta_4\} \quad \mathcal{G}_2^3 = \{\delta_6, \delta_8\}$$

$$\mathcal{G}_3^1 = \{\delta_1, \delta_4, \delta_8, \delta_9\} \quad \mathcal{G}_3^2 = \{\delta_1, \delta_8\} \quad \mathcal{G}_3^3 = \{\delta_4, \delta_9\}$$

III - Expériences - Détection visages





Évolution de $|\mathcal{F}_t|$ Après zoom

Diminution du nombre de variables, mais $\mathcal{F}_t$ non stabilisé

Faux positif nul. Erreur globable 1.5% pour un vote de 10 experts par classe utilisant chacun 50 variables.

III - Expériences - Détection visages

IV - Conclusion

- Formalisation d’une méthode de sélection de variables
- Diminution de taux d’erreur intéressant, facilité d’interprétation pour la sélection de variables
- Formalisation d’une méthode de construction d’un nouvel espace de variables (variables binaires, ternaires)
- Taux d’erreurs améliorés par rapport au processus sans saut
- Réduction de la complexité des données
- Approche généralisable à d’autres problèmes de classification

IV - Conclusion - Limite du modèle

- Grande quantité d'exemples nécessaires
- Temps d'apprentissage long et couteux en mémoire, même si la phase de test est très rapide
- Performances décevantes sur certaines bases de données d'images (chiffres manuscrits): pas d'amélioration notable avec le processus de sauts

IV - Conclusion - Perspectives

- Implémenter une approche ou les détecteurs de bords sont invariants par translation (classification d'images)
- Appliquer algorithme de sélection de variables sur des bases de données génétiques (Microarray)
- Étudier via le processus de diffusion avec sauts, la formation d'une syntaxe (du moins la formation des mots) pour des problèmes de classification de textes