
**Une méthode d'ensemble
en apprentissage non supervisée
quand on ne connaît rien
sur la performance des experts ?**

Antoine Cornuéjols & Christine Martin

AgroParisTech – INRA MIA 518

antoine.cornuejols,christine.martin@agroparistech.fr

Plan

1. Introduction : le problème

2. Approches conventionnelles

1. Clustering
2. Fonctions d'évaluation

3. Approche proposée

1. Principe
2. Algorithme de sélection des bonnes fonctions d'évaluation

4. Expériences

1. Protocoles
2. Résultats
3. La combinaison des résultats

5. Conclusions

Introduction

Le problème

Découverte d'une classe d'intérêt

- Illustrations
 - Identification de **fraudeurs**
 - **Gènes** activés dans une condition environnementale
 - **Protéines** interagissant avec un médicament

Découverte d'une classe d'intérêt

- Apprentissage **non supervisé**

$$\mathcal{S} = \{\mathbf{x}_1, \dots, \mathbf{x}_m\}$$

- à **2** (?) classes

$$\mathcal{S} = \{\mathbf{x}_1, \dots, \mathbf{x}_{m^+}, \mathbf{x}_{m^++1}, \dots, \mathbf{x}_m\}$$

$\mathbf{P}_{\mathcal{X}}^+$

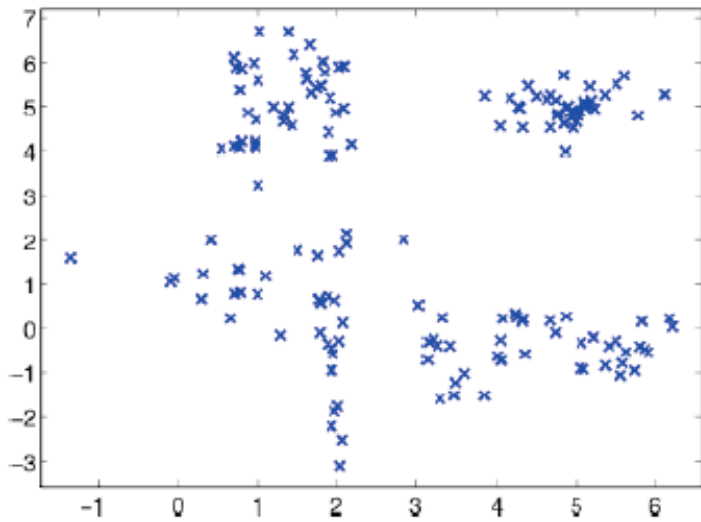


$\mathbf{P}_{\mathcal{X}}^-$



Approches conventionnelles

Clustering



Clustering

- Questions

1. « Bonne » distance ?
2. 2 classes ou N classes ?
 - Problème de multi-modalité
3. Si déséquilibré ?

Fonctions d'évaluation

- Fonction

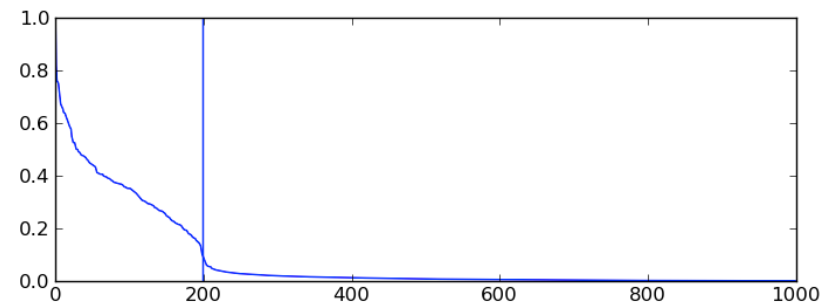
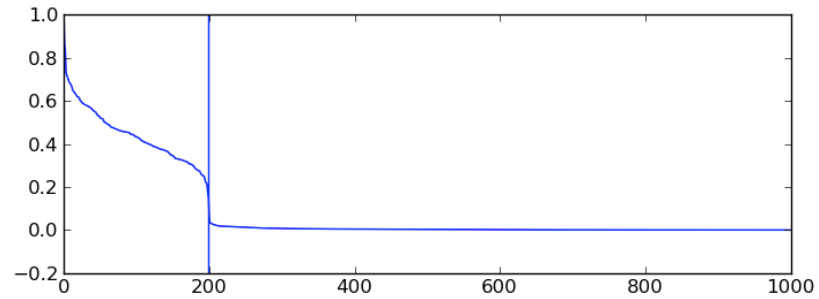
$$f_i : \mathcal{X} \rightarrow \mathbb{R}$$

- Permettant de classer les « objets »

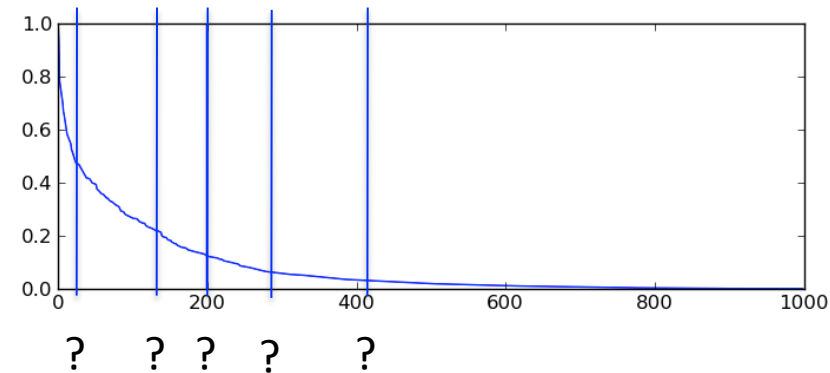
x₂₃
x₄₅
x₁₇
x₀₅
x₇₃
x₅₂
x₁₈
x₀₄
x₂₆
x₀₁
...
x₂₃₄
x₈₄₅
x₇₅₆

Fonctions d'évaluation

- Tri des exemples
 - Filter method



Seuil ?



Fonctions d'évaluation

- Questions
 - A-t-on une **bonne fonction** ?
 - Existe-t-il un **seuil** évident ?

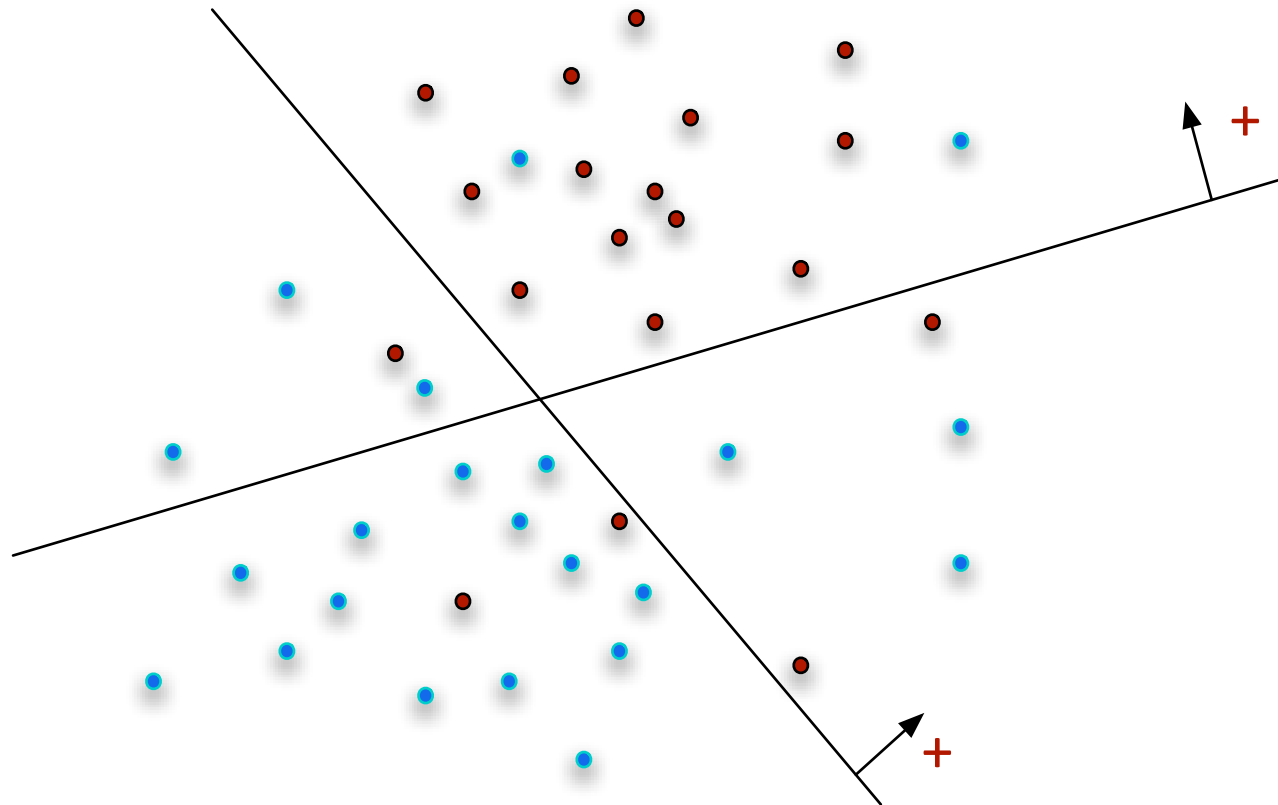
Approche proposée

Une méthode d'ensemble

- Un **ensemble** \mathcal{F} de fonctions d'évaluation
- On mesure la **corrélation des tris**
 - Sur S
 - Sur des échantillons aléatoires
- On **retient les fonctions**
 - Sur-corrélées sur S
 - D'accord sur S
 - Pas (peu) en général

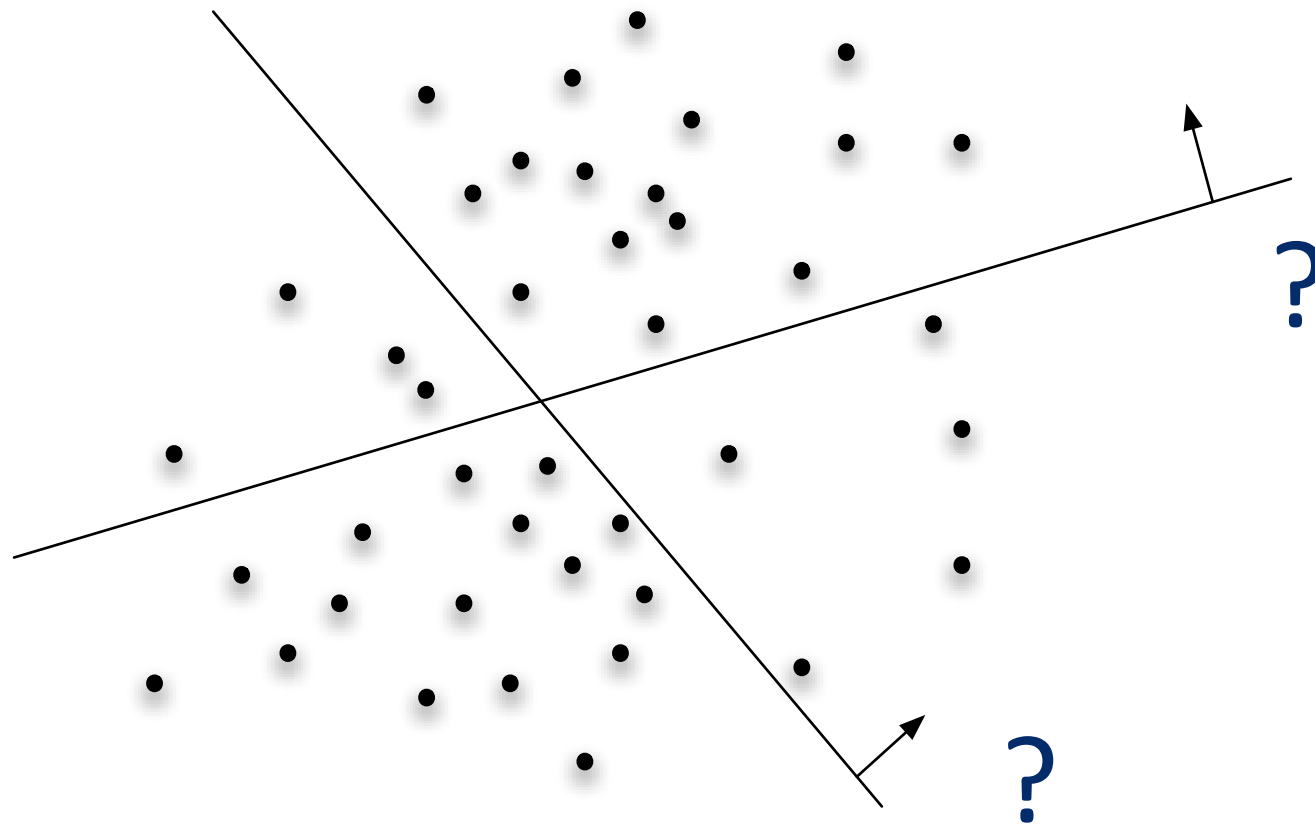
Méthode d'ensemble supervisée

- Le **boosting**

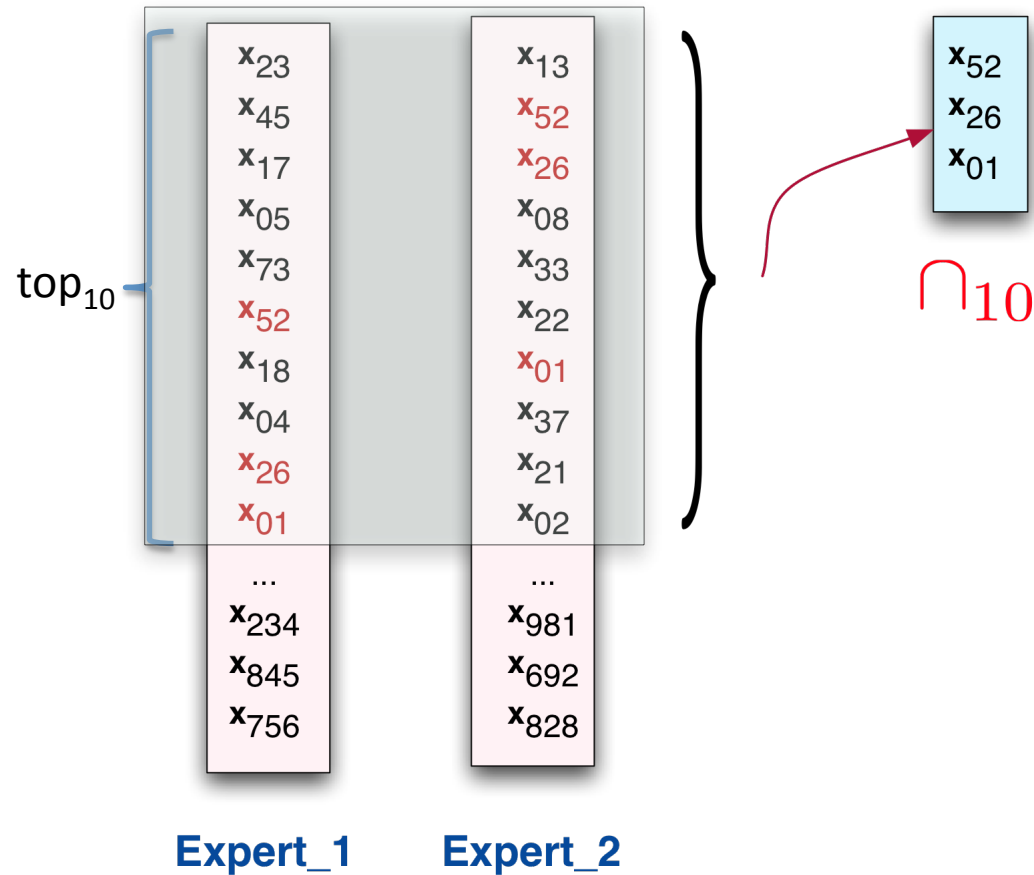


Méthode d'ensemble **non** supervisée

- Nouveau problème

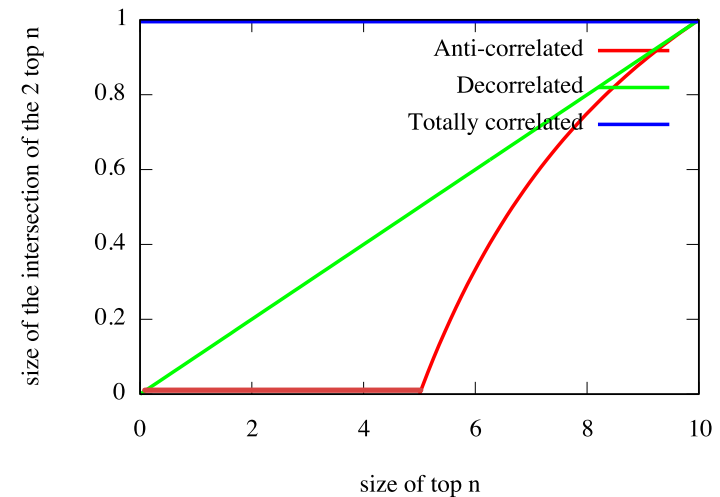


Mesure de corrélation de tris

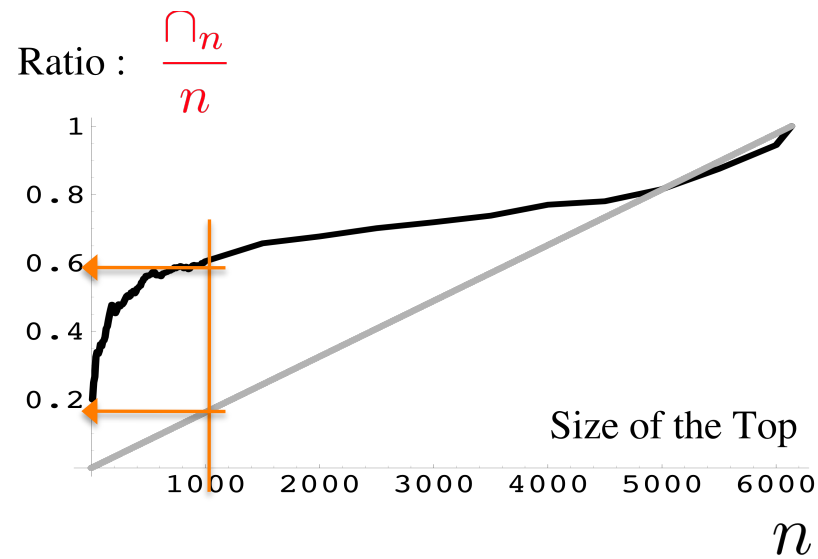


Détection de « nouveautés »

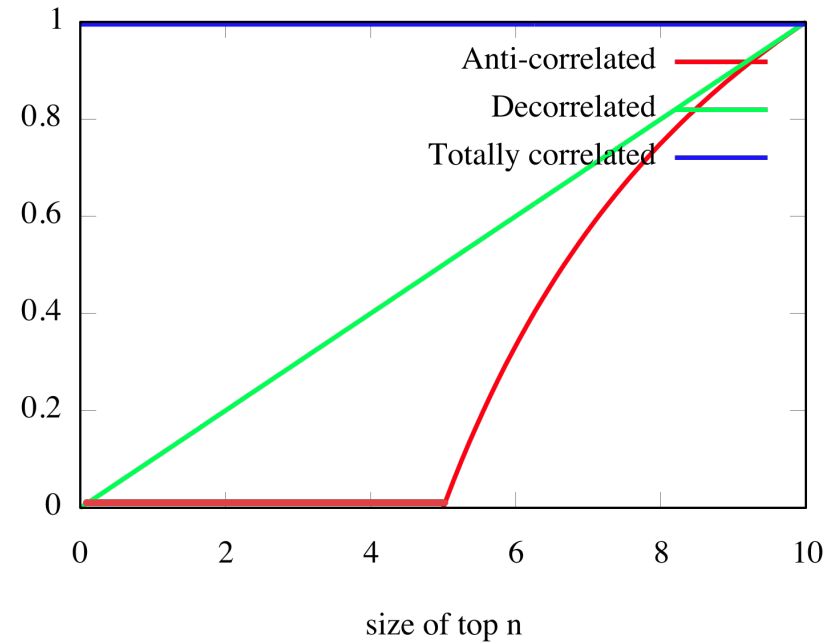
- **Corrélation** entre fonctions d'évaluation
 - **Quelle information** est fournie par la valeur ou le classement d'un exemple **par une fonction d'évaluation** sur la valeur ou le classement d'**une autre fonction** ?
 - Spearman / Kendall
 - **Notre approche**
 - Intersection des top_ n des deux fonctions étudiées sur l'échantillon S .



Mesurer la corrélation

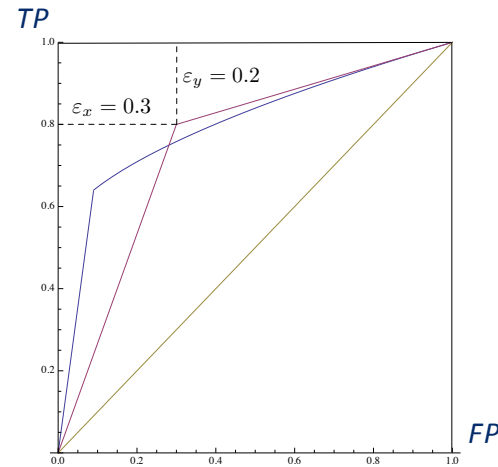
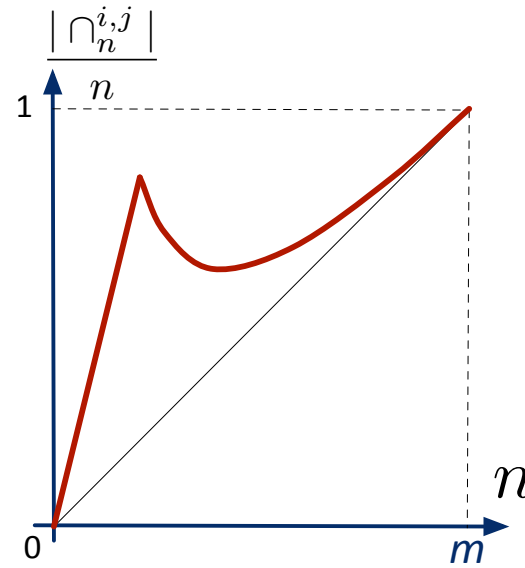
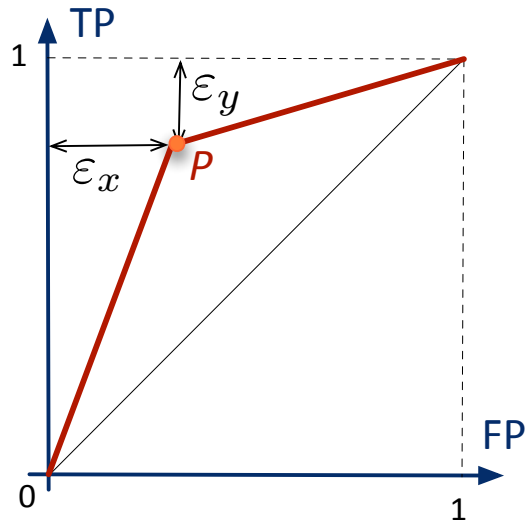


Une courbe réelle

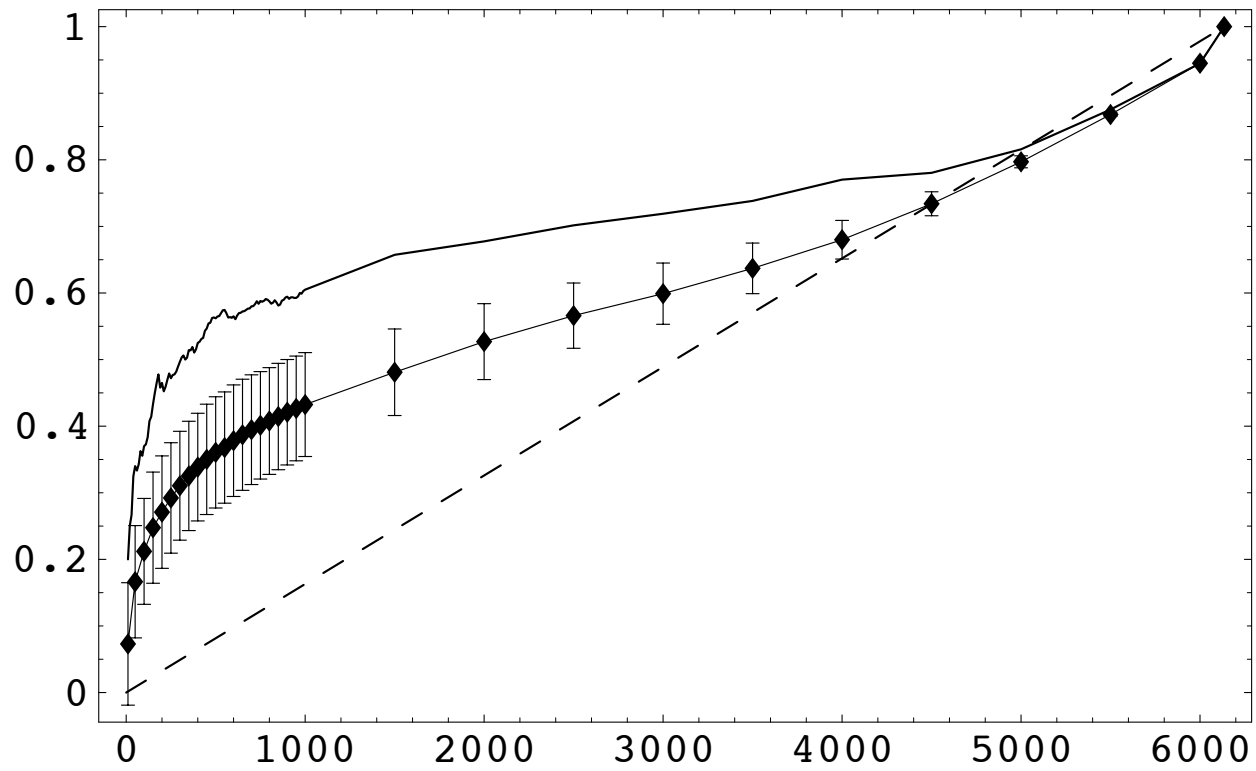


Enveloppe de courbes possibles

Étude théorique



Mesure de sur-corrélation



L'algorithme de sélection

Algorithme 1: Sélection de fonctions de base pertinentes

Entrées : La base d'exemples \mathcal{S}

L'ensemble \mathcal{F} de fonctions d'évaluation de base

Sorties : Un sous-ensemble $\mathcal{F}'' \in \mathcal{F}$ de fonctions de base

Génération de N échantillons "aléatoires" \mathcal{S}_0 ;

pour tous les couples de fonctions d'évaluation $(f_i, f_j)_{(i \neq j)} \in \mathcal{F}$ **faire**

Calculer la surcorrélacion de (f_i, f_j) sur \mathcal{S} par rapport à la corrélation
 moyenne sur les échantillons \mathcal{S}_0

fin pour tous

Sélectionner les fonctions d'évaluation $f_i \in \mathcal{F}$ de surcorrélacion \geq

seuil_min_surcor : soit \mathcal{F}'

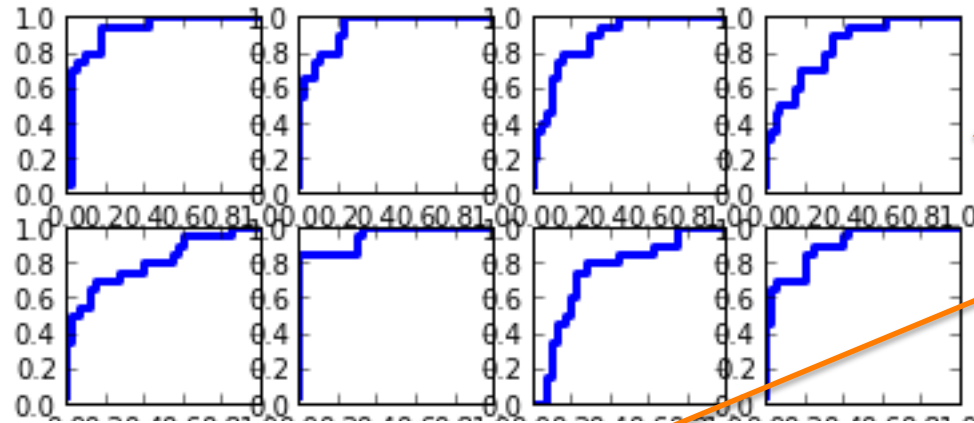
Initialisation : $\mathcal{F}'' = \emptyset$

pour tous les $f_i \in \mathcal{F}'$ **faire**

si $\sum_{j \neq i} \text{surcorr}(f_i, f_j) \geq \text{seuil}$ **alors**
 Mettre f_i dans \mathcal{F}''

fin pour tous

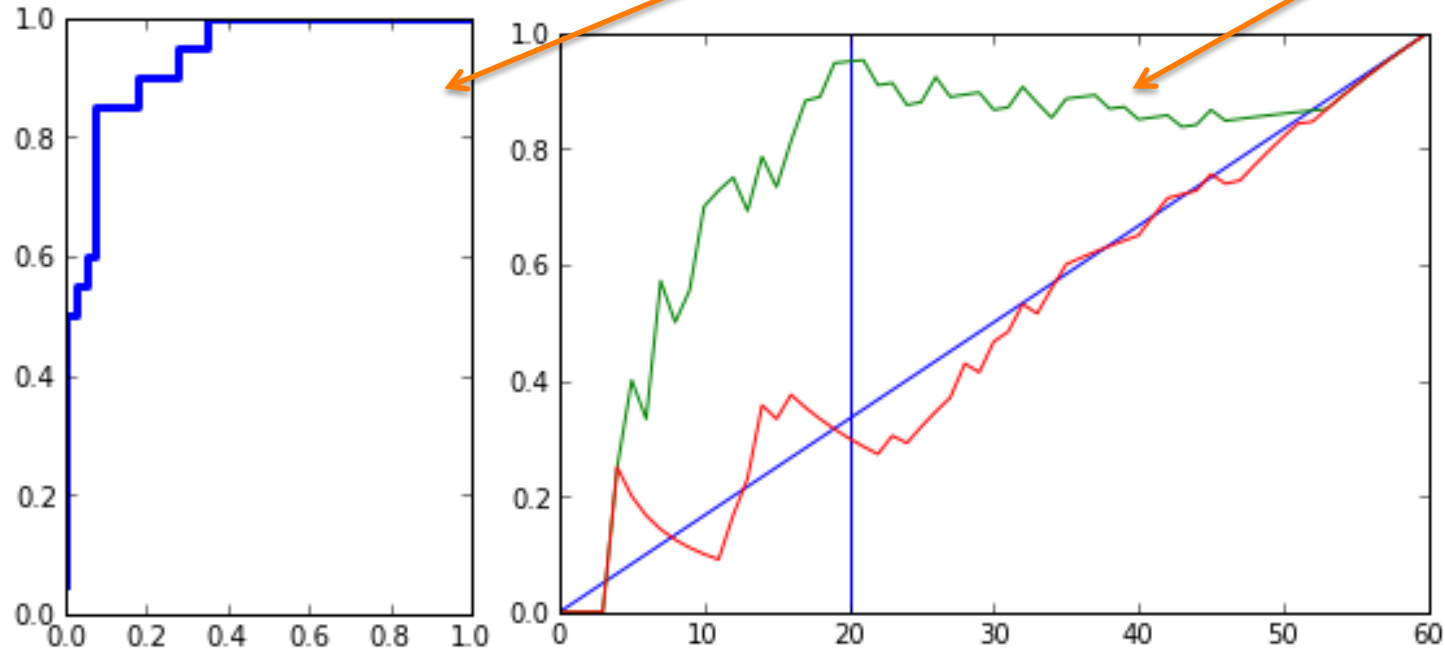
Illustration



Courbes ROC des fonctions sélectionnées

Courbe ROC de la combinaison

Corrélation des deux 1ères fonctions



Expériences

Protocole expérimental

- **320 exemples** dont
 - **40** (soit 1/8) ; **80** (soit 1/4) ; **120** (soit 1/3) d'exemples '+'
 - Dans \mathbb{R}^{20}
- 2 gaussiennes : \mathbf{P}_x^+ et \mathbf{P}_x^-
 - $|\mu_+ - \mu_-|_2 = 3$
 - $\sigma = 1,5$ ou $2,5$ ou $3,5$ ou $4,5$ (taux de bruit)
- **45 fonctions** d'évaluation
 - 22 positivement alignées
 - 22 négativement alignées
 - 1 fonction aléatoire

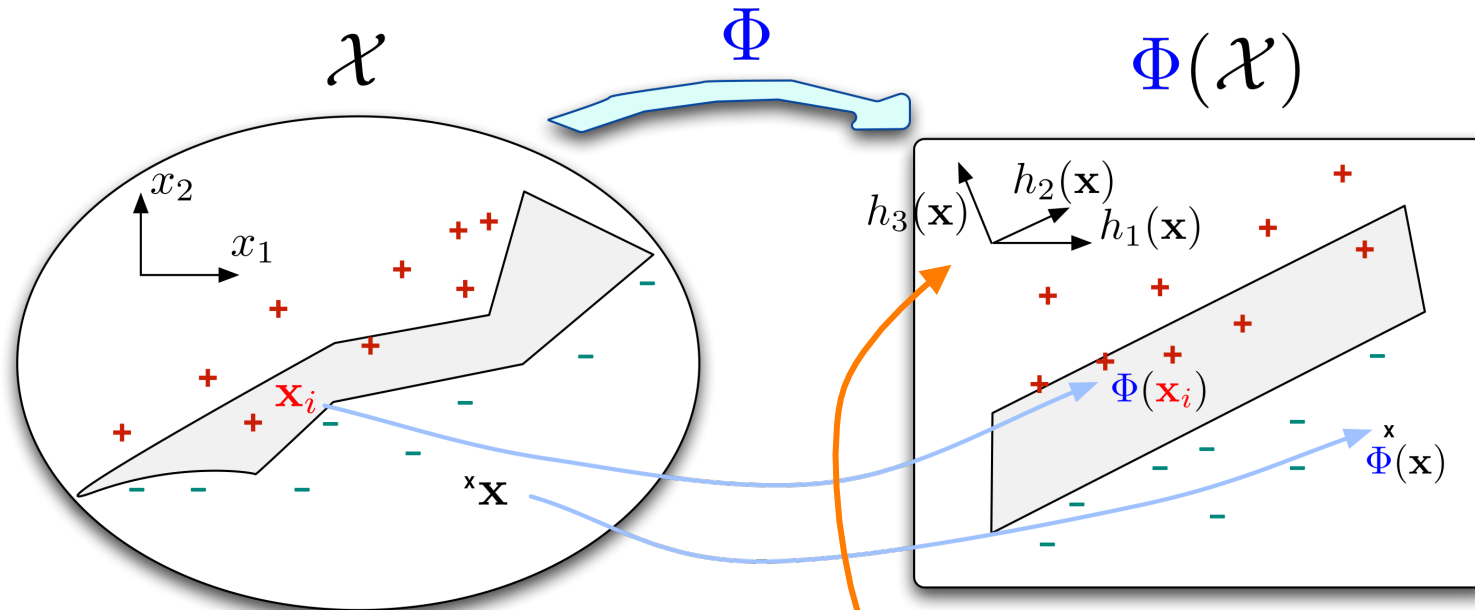
Résultats expérimentaux

σ	$\frac{m^+}{m}$	<i>Before selection</i>		<i>After selection</i>			AUC comb
		auc_m	auc^M	auc_m	auc^M	$\overline{\text{auc}}$	
1.5	$\frac{40}{320}$	0 ± 0	1 ± 0	0.92 ± 0.03	1 ± 0	0.98 ± 0.01	1 ± 0
	$\frac{80}{320}$	0 ± 0	1 ± 0	0.87 ± 0.06	1 ± 0	0.97 ± 0.01	1 ± 0
	$\frac{120}{320}$	0 ± 0	1 ± 0	0.84 ± 0.07	1 ± 0	0.95 ± 0.01	1 ± 0
	$\frac{320}{320}$	0 ± 0	1 ± 0	0.84 ± 0.07	1 ± 0	0.95 ± 0.01	1 ± 0
2.5	$\frac{40}{320}$	0.02 ± 0.01	0.98 ± 0.01	0.94 ± 0.03	0.98 ± 0.00	0.96 ± 0.02	0.98 ± 0.01
	$\frac{80}{320}$	0.03 ± 0.01	0.98 ± 0.01	0.85 ± 0.05	0.98 ± 0.01	0.91 ± 0.02	0.97 ± 0.01
	$\frac{320}{120}$	0.03 ± 0.01	0.98 ± 0.01	0.76 ± 0.03	0.98 ± 0.01	0.88 ± 0.02	0.97 ± 0.01
	$\frac{320}{160}$	0.03 ± 0.01	0.98 ± 0.01	0.73 ± 0.04	0.97 ± 0.01	0.85 ± 0.02	0.95 ± 0.01
	$\frac{320}{320}$	0.03 ± 0.01	0.98 ± 0.01	0.73 ± 0.04	0.97 ± 0.01	0.85 ± 0.02	0.95 ± 0.01
3.5	$\frac{40}{320}$	0.09 ± 0.02	0.91 ± 0.02	0.75 ± 0.06	0.90 ± 0.03	0.83 ± 0.01	0.90 ± 0.03
	$\frac{80}{320}$	0.09 ± 0.02	0.92 ± 0.02	0.65 ± 0.05	0.92 ± 0.02	0.79 ± 0.02	0.90 ± 0.02
	$\frac{320}{120}$	0.09 ± 0.02	0.91 ± 0.01	0.64 ± 0.04	0.91 ± 0.01	0.77 ± 0.02	0.89 ± 0.02
	$\frac{320}{160}$	0.10 ± 0.01	0.91 ± 0.02	0.63 ± 0.03	0.91 ± 0.02	0.76 ± 0.02	0.88 ± 0.02
	$\frac{320}{320}$	0.10 ± 0.01	0.91 ± 0.02	0.63 ± 0.03	0.91 ± 0.02	0.76 ± 0.02	0.88 ± 0.02
4.5	$\frac{40}{320}$	0.13 ± 0.02	0.86 ± 0.02	0.67 ± 0.03	0.86 ± 0.02	0.76 ± 0.02	0.86 ± 0.02
	$\frac{80}{320}$	0.15 ± 0.02	0.85 ± 0.02	0.65 ± 0.03	0.84 ± 0.03	0.75 ± 0.02	0.84 ± 0.03
	$\frac{320}{120}$	0.15 ± 0.02	0.84 ± 0.02	0.62 ± 0.06	0.84 ± 0.02	0.73 ± 0.03	0.84 ± 0.02
	$\frac{320}{160}$	0.15 ± 0.01	0.85 ± 0.01	0.61 ± 0.03	0.85 ± 0.01	0.72 ± 0.02	0.83 ± 0.03

Table 1: Experimental results in function of the noise parameter σ and the proportion of the class ‘+’.

La combinaison des fonctions d'évaluation

Boosting et redescription

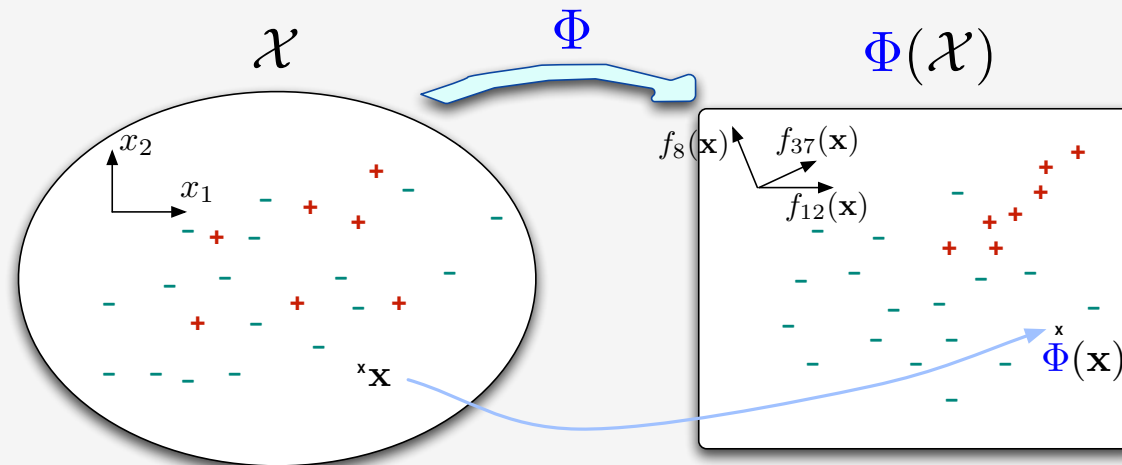


$$H(\mathbf{x}) = \text{sign} \left\{ \sum_{i=1}^T \alpha_i h_i(\mathbf{x}) \right\}$$

- Construction **itérative** de l'espace de redescription

Détection de nouveautés

- Algorithme itératif de combinaison de fonctions d'évaluation
 - De plus en plus discriminatives



- 📄 A. Cornuéjols, Ch. Martin (2011). « Unsupervised Object Ranking Using not even Weak Experts ». *Int. Conf. on Neural Information Processing (ICONIP-2011)*.
- 📄 Ch. Martin & A. Cornuéjols (2012). « Détection non supervisée d'une sous-population par méthode d'ensemble et changement de représentation itératif ». *Conf. Extraction et Gestion des Connaissances, (EGC-2012)*.
- 📄 A. Cornuéjols, Ch. Martin (2012). « Une nouvelle méthode de combinaison d'outils d'identification non supervisés ». *Atelier Prospectom-2012, Grenoble, 29-30 Nov. 2012*.
- 📄 A. Cornuéjols, Ch. Martin (AAFD-2014). « Une méthode d'ensemble en apprentissage non supervisé quand on ne connaît rien sur la performance des experts ? ». *AAFD-2014, 29-30 avril 2014, LIPN*.

Conclusions

Conclusions

- En apprentissage non supervisé (à 2 classes)
 - 1^{ère} **méthode d'ensemble** ne **présupposant pas** des « experts » bons
- Nouveau **critère de sélection**
 - La **sur-corrélation** des tris
- Étude **théorique**
 - On peut **améliorer** la **précision** et le **rappel** autant que l'on veut (peut)
- Étude **empirique**
 - **Bons résultats** confirmant le bien-fondé de l'approche
- Perspectives pour **améliorer la combinaison**