

Identification de paramètres et optimisation

Cours de Master 2 STIM

2014-2015

Sébastien Adam

10 février 2016

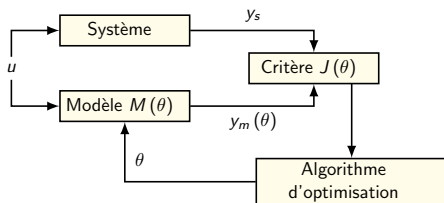
Plan du Cours

- 1 Présentation de l'enseign(ant)ement
- 2 Introduction
- 3 Méthodes des moindres carrés
- 4 Méthodes de descente locale
- 5 Réseaux de neurones
- 6 Méthodes itératives globales
- 7 Optimisation multi-objectif
- 8 Optimisation multi-objectif**

Optimisation multi-objectif (1)

Contexte : la modélisation paramétrique

- Objectif : identifier les paramètres θ du modèle d'un système qui maximise un critère scalaire de ressemblance.



- K paramètres
- N observations du système

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_K \end{pmatrix} \quad y_S = \begin{pmatrix} Y_{S1} \\ Y_{S2} \\ \vdots \\ Y_{SN} \end{pmatrix} \quad y_M = \begin{pmatrix} y_{M1} \\ y_{M2} \\ \vdots \\ y_{MN} \end{pmatrix}$$

Vers des critères multiples

- Beaucoup de problèmes réels ont des objectifs multiples
 $\rightarrow J(\theta) = (J_1(\theta), J_2(\theta), \dots, J_n(\theta))^T$
- Les $J_i(\theta)$ sont souvent antagonistes

Optimisation multi-objectif (2)

Exemple 1 : le sac à dos

- Exemple : comment remplir son sac à dos avec les objets suivants, en respectant un poids maximal et en maximisant le profit
 - ▶ 1 appareil photo → profit : 5, poid 750
 - ▶ 1 couteau suisse → profit : 7, poid 300
 - ▶ 1 bouteille thermos → profit : 8, poid 1500
 - ▶ 1 bouquin → profit : 3, poid 1000
- Remarques :
 - ▶ Problèmes d'optimisation combinatoire très étudié depuis 1950
 - ▶ Très nombreux algos proposés : glouton, métaheuristique, prog dynamique, B&B, Hybrides...
 - ▶ Très nombreux problèmes dérivés : LKP (variables continues), variables entières (occurrences multiples), multidimensionnel (sur la contraintes), quadratique (gains co-occurrences), choix multiples (classes d'objets)
 - ▶ Version multiobjectifs

Optimisation multi-objectif (2)

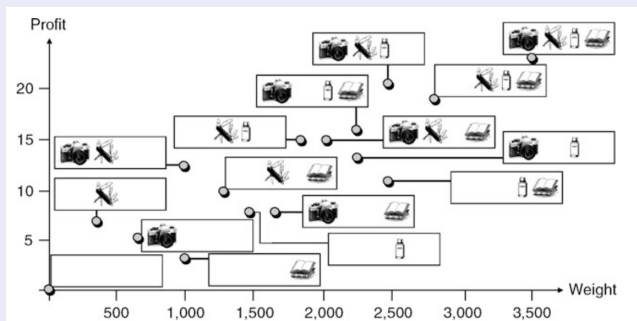
Exemple 1 : le sac à dos

- Comment remplir son sac à dos avec les objets suivants :
 - ▶ 1 appareil photo → profit : 5, poids 750
 - ▶ 1 couteau suisse → profit : 7, poids 300
 - ▶ 1 bouteille thermos → profit : 8, poids 1500
 - ▶ 1 bouquin → profit : 3, poids 1000
- 16 configurations

Optimisation multi-objectif (2)

Exemple 1 : le sac à dos

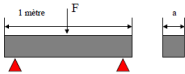
- Comment remplir son sac à dos avec les objets suivants :
 - ▶ 1 appareil photo → profit : 5, poids 750
 - ▶ 1 couteau suisse → profit : 7, poids 300
 - ▶ 1 bouteille thermos → profit : 8, poids 1500
 - ▶ 1 bouquin → profit : 3, poids 1000
- 16 configurations



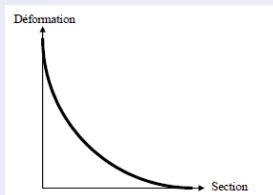
Optimisation multi-objectif (3)

Exemple 2 : mécanique

- On cherche à déterminer la section d'une poutre telle que :
 - ▶ Son poid (donc sa section) soit minimal
 - ▶ Elle se déforme le moins possible

$S(a) = a^2$ $d(a) = 1000 + \frac{1 \cdot 10^{-2}}{192 + 2 \cdot 10^5 + \frac{a^4}{12}}$ $a \leq 0.1$	
---	---

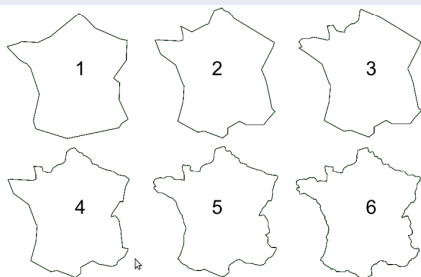
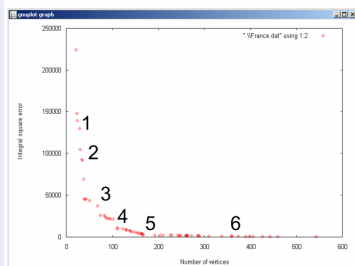
- Les deux objectifs sont antagonistes : on cherche un compromis



Optimisation multi-objectif (4)

Exemple 3 : approximation de courbes

- On cherche à approximer une courbe par une séquence de points telle que :
 - Le nombre de point soit minimal
 - L'erreur d'approximation soit minimale
- Les deux objectifs sont antagonistes : on cherche un compromis

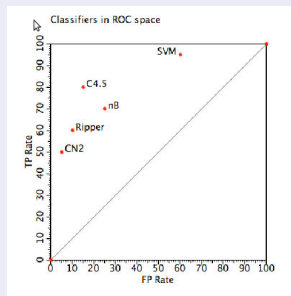


Optimisation multi-objectif (5)

Exemple 4 : classification ROC

- On cherche à optimiser un classifieur à deux classes tel que
 - Le taux de vrais positifs (TPR) soit maximal
 - Le taux de faux positifs (FPR) soit minimal
- Les deux objectifs sont antagonistes : on cherche un compromis

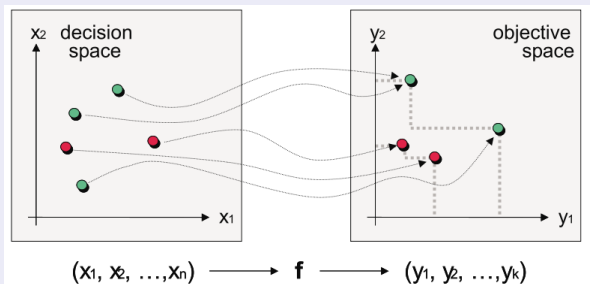
	Predicted positive	Predicted negative	
Positive examples	TP	FN	Pos
Negative examples	FP	TN	Neg
	PPos	PNeg	N



Optimisation multi-objectif (6)

Formulation générale du problème d'optimisation

$$\left. \begin{array}{ll} \text{Minimiser} & f_k(\vec{x}) \\ \text{sous contrainte de} & \begin{array}{l} g_j(\vec{x}) \geq 0 \\ h_l(\vec{x}) = 0 \\ x_i^L \leq x_i \leq x_i^U \end{array} \end{array} \right\} \begin{array}{l} k \in [1, K] \\ j \in [1, J] \\ l \in [1, L] \\ i \in [1, N] \end{array}$$

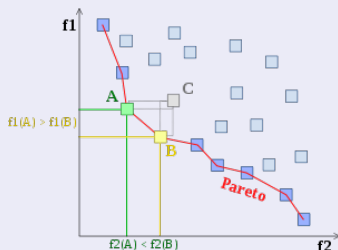


Optimisation multi-objectif (7)

Conséquence du nombre multiple d'objectifs

- Il n'existe en général plus UNE meilleure solution
- Il y a des solutions meilleures que d'autres (A vs C)
- Il y a des solutions équivalentes (A vs B)

→ Perte de la relation d'ordre total

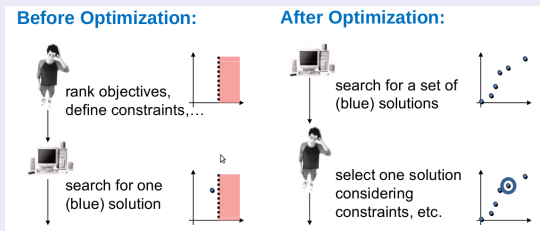


La notion de dominance [Vilfredo Pareto]

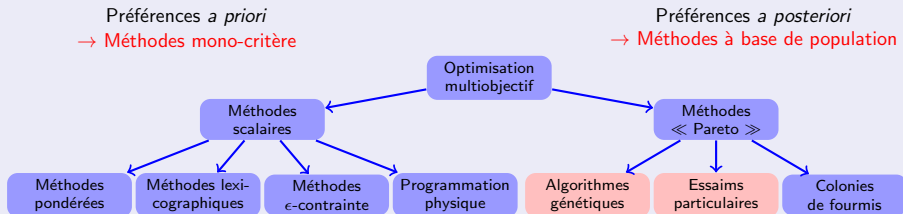
Definition Une solution \vec{x} domine une autre solution \vec{y} si et seulement si $\forall k \in [1, K], f_k(\vec{x}) \leq f_k(\vec{y})$ et si $\exists k \in [1, K] / f_k(\vec{x}) < f_k(\vec{y})$. Une telle relation est notée $\vec{x} \prec \vec{y}$

Optimisation multi-objectif (8)

Deux options pour résoudre de tels problèmes



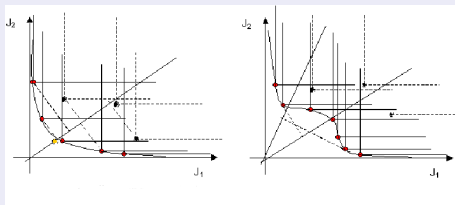
Deux familles d'approches



Optimisation multi-objectif (9)

Principes des méthode pondérée

- Approche la plus classique : facile à mettre en œuvre
- Repose sur la minimisation d'une combinaison des objectifs
- Exemple : $J(\theta) = \sum_{i=1}^k w_i J_i(\theta)$
- $w_i \geq 0$: importance relative des objectifs du problème.



- Peut fournir un ensemble de compromis mais ...
- Coûteux et inefficace si le Front de Pareto est concave
- Variantes : exponentielles, Tchebycheff, ...

Optimisation multi-objectif (10)

Principes des méthodes lexicographique

- Objectifs triés par ordre d'importance
- Optimisations scalaires successives sous contraintes

$$\min_{x \in X} F_i(x)$$

$$\text{subject to } F_j(x) = F_j(x_j^*)$$

$$j = 1, 2, \dots, i-1, i > 1$$

$$i = 1, 2, \dots, k$$

- Problème 1 : minimisation du premier critère emmène l'algorithme vers une zone restreinte, d'où les autres critères ne peuvent plus sortir.
- Problème 2 : critères sont très souvent liés, et surtout, antagonistes

Optimisation multi-objectif (11)

Principes des méthodes ε -constraint

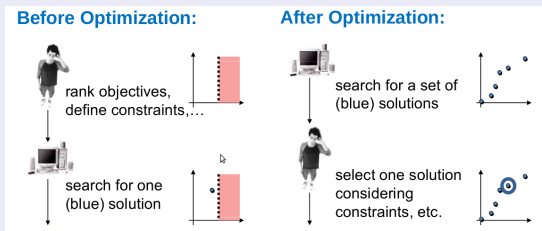
- Principe : minimiser uniquement l'objectif le plus important (ex J_i).
- Les autres objectifs J_k avec $k \neq i$ sont exprimés sous forme de contraintes telles que :

$$J_k(x) \leq \varepsilon_k \quad (1)$$

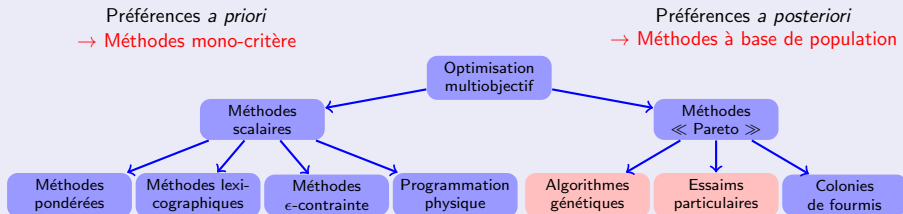
- En faisant varier systématiquement les ε_k , on obtient un ensemble de solution du front de Pareto, en atteignant les concavités.
- Par contre, un mauvais choix d' ε_k peut entraîner une formulation non faisable.

Optimisation multi-objectif (12)

Deux options pour résoudre de tels problèmes



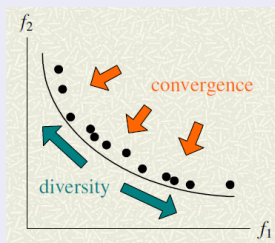
Deux familles d'approches



Optimisation multi-objectif (13)

Principes des méthodes "Pareto"

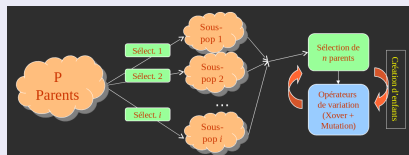
- Injecter les préférences a posteriori, au regard des compromis possibles
- Générer un ensemble de solutions (compromis) non dominées → Les méthodes à base de population sont une réponse naturelle !!!
- Objectifs pour l'algorithmes d'optimisation
 - ▶ Se rapprocher le plus possible du Front de Pareto du problème
 - ▶ Couvrir au mieux l'ensemble du front (diversité + spreading)
- L'optimisation est elle même un problème à objectifs multiples



Optimisation multi-objectif (14)

Première approche à base de population : VEGA (Schaefer[85])

- i objectifs, P individus.
- Sélection de P/i individus pour chaque critère $\rightarrow i$ sous-populations contenant les meilleurs individus pour chacun des critères
- Sélection interclassée des individus

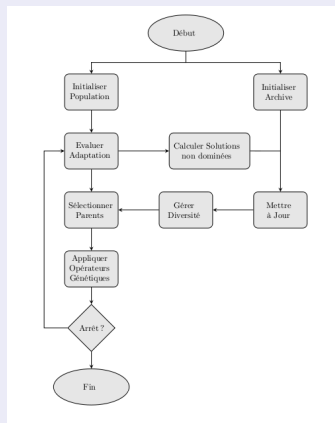


- Très facile à implémenter, variantes en modifiant les opérateurs de croisement pour inciter les mix de sous populations
- Individus très bons dans un seul domaine privilégiés, au détriment d'individus bons partout \rightarrow Approche non Pareto

Optimisation multi-objectif (15)

Schéma général des méthodes "Pareto"

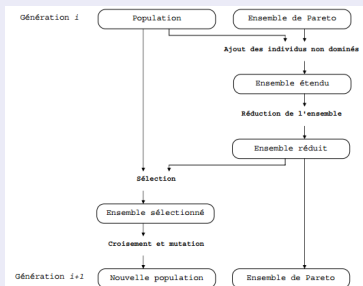
- Utilisation de la notion de rang de dominance
- Utilisation d'une archive conservant les solutions non dominées
- Gestion de la diversité



Optimisation multi-objectif (16)

Exemple de SPEA (Zitzler et Thiele en 1998)

- Utilisation d'une archive contenant les individus non dominés

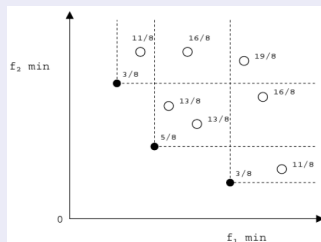


- A chaque itération :
 - ▶ Sélection dans (archive+population) sur la base d'un score
 - ▶ Enfants créés puis évalués, enfants non dominés → archive
 - ▶ Éléments dominés de l'archive sortis
 - ▶ Si la taille de l'archive est insuffisante : clustering

Optimisation multi-objectif (16)

Exemple de SPEA (Zitzler et Thiele en 1998)

- Fitness d'une solution dominée : somme des "forces" des individus de l'archive qui la dominent +1
- Force $S = \text{nb d'individus dominés} / (\text{nb d'individus de la pop} + 1)$



- Notion de Force + Clustering → Bonne distribution des solutions
- Formation de niches

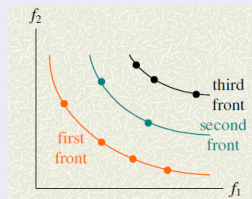
Optimisation multi-objectif (17)

Exemple de NSGA-II

- Proposé par Deb en 2001
- Version améliorée d'un premier algorithme (NSGA)
- Utilisé comme référence dans beaucoup d'articles
- Originalités :
 - ▶ Fitness basé sur le rang de non-dominance
 - ▶ Méthode de Crowding-distance pour préserver la diversité

Attribution du fitness

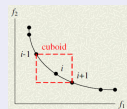
- Les solutions non dominées ont le fitness le plus important
- On procède par suppressions successives



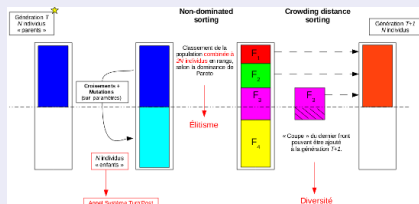
Optimisation multi-objectif (18)

Préservation de la diversité

- distance de crowding = taille du cuboid



Algorithme NSGA-II



Algorithm 1 NSGA-II algorithm

```

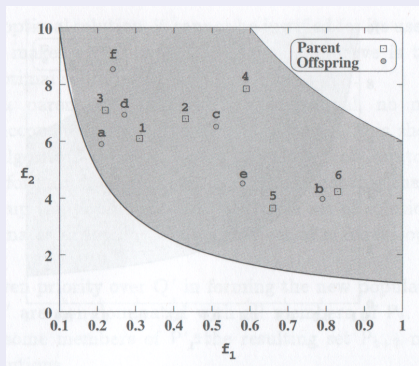
 $P_0 \leftarrow \text{pop-init}()$ 
 $Q_0 \leftarrow \text{make-new-pop}(P_0)$ 
 $t \leftarrow 0$ 
while  $t < M$  do
   $R_t \leftarrow P_t \cup Q_t$ 
   $\mathcal{F} \leftarrow \text{non-dominated-sort}(R_t)$ 
   $P_{t+1} \leftarrow \emptyset$ 
   $i \leftarrow 0$ 
  while  $|P_{t+1}| + |\mathcal{F}_i| \leq N$  do
     $P_{t+1} \leftarrow P_{t+1} \cup \mathcal{F}_i$ 
    crowding-distance-assignment( $\mathcal{F}_i$ )
     $i \leftarrow i + 1$ 
  end while
  Sort( $\mathcal{F}_i, <_n$ )
   $P_{t+1} \leftarrow P_{t+1} \cup \mathcal{F}_i[1 : (N - |P_{t+1}|)]$ 
   $Q_{t+1} \leftarrow \text{make-new-pop}(P_{t+1})$ 
   $t \leftarrow t + 1$ 
end while

```

Optimisation multi-objectif (19)

Illustration

Parent population, P_t					Offspring population, Q_t				
Solution	x_1	x_2	f_1	f_2	Solution	x_1	x_2	f_1	f_2
1	0.31	0.89	0.31	6.10	a	0.21	0.24	0.21	5.90
2	0.43	1.92	0.43	6.79	b	0.79	2.14	0.79	3.97
3	0.22	0.56	0.22	7.09	c	0.51	2.32	0.51	6.51
4	0.59	3.63	0.59	7.85	d	0.27	0.87	0.27	6.93
5	0.66	1.41	0.66	3.65	e	0.58	1.62	0.58	4.52
6	0.83	2.51	0.83	4.23	f	0.24	1.05	0.24	8.54



Optimisation multi-objectif (20)

Très nombreuses autres méthodes

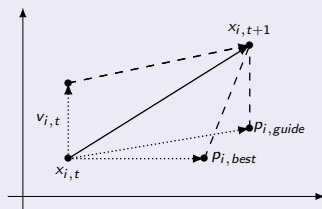
- PAES (Pareto Archived Evolution Strategy), Knowles 2000
 - ▶ Pas d'utilisation de population : (1,1)-ES
 - ▶ Utilisation d'une archive contenant l'approximation courante du front
 - ▶ Technique de crowding basée sur un découpage en hypercubes
 - ▶ Principales étapes
 - ① Génération aléatoire d'une solution c et ajout dans l'archive
 - ② Production et évaluation d'une solution m par mutation de c
 - ③ Si c domine m écarter m
 - ④ Si m domine c remplacer c par m
 - ⑤ Si m est dominé par une solution de l'archive écarter m
 - ⑥ Sinon Appliquer Test($c, m, \text{archive}$)
 - ▶ Test($c, m, \text{archive}$) :
 - ① Si l'archive n'est pas pleine : ajout de m à l'archive
 - ② Sinon, si m est dans une région moins encombrée qu'une solution x , changer x par m
 - ③ Remplacer c par m si elle est dans une région moins encombrée

Optimisation multi-objectif (21)

Rappel essais particulaires

- Population de particules

$$\begin{aligned}
 v_{i,t+1} &= \omega \cdot r_0 \cdot v_{i,t} + \\
 &\quad c_1 \cdot r_1 \cdot (p_{i,best} - x_{i,t}) + \\
 &\quad c_2 \cdot r_2 \cdot (p_{i,guides} - x_{i,t}) \\
 x_{i,t+1} &= x_{i,t} + \chi(v_{i,t+1})
 \end{aligned}$$



Adaptation aux problèmes multiobjectifs

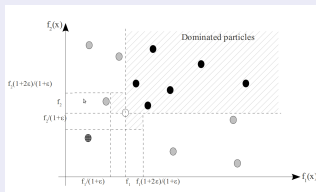
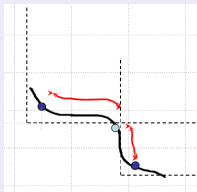
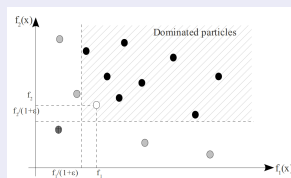
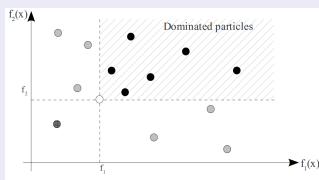
- Deux difficultés :
 - ▶ Gestion d'une archive → Problèmes de taille et de diversité
 - ▶ Gestion du guide et de la mémoire de la particule ($p_{i,best}$ et $p_{i,guides}$)
- Contributions
 - ▶ Nouvelle dominance inspirée de l' ϵ -dominance
 - ▶ Sélection stochastique du guide guidée par la diversité

Optimisation multi-objectif (22)

Gestion de l'archive

- Nouvelle définition de la dominance incluant la diversité

Nouvelle dominance

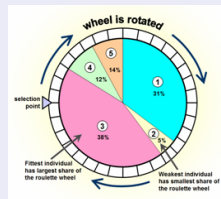
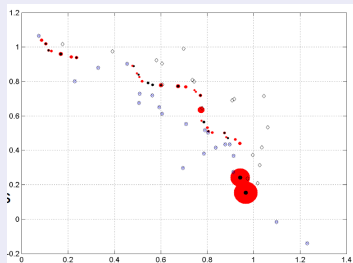


Optimisation multi-objectif (23)

Gestion des guides

- Sélection probabiliste du guide : $P(p_i) = \frac{1}{densite}$
- Mémorisation d'un guide pour n itérations

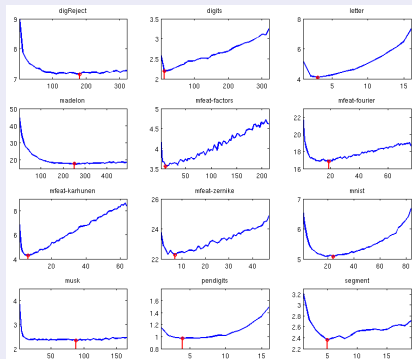
Sélection probabiliste du guide



Optimisation multi-objectif : application (24)

Les hyperparamètres

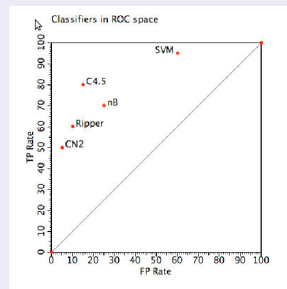
- Un problème crucial de la sélection de modèle [Bernard]



- Une littérature très « scalaire »

Espace ROC

- Problèmes à 2 classes mal définis (médical, fraude, biometrie)
 - ▶ Effectifs non balancés
 - ▶ Peu de données
 - ▶ Coût inconnus



Optimisation multi-objectif : application (25)

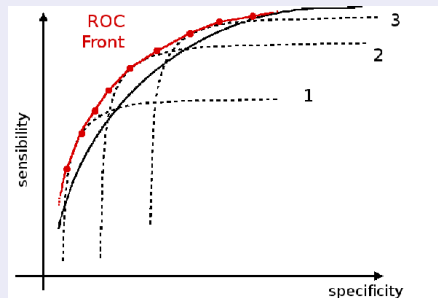
Approches courantes

- ϵ -contrainte + Neyman-Pearson
→ Quelles bornes ?
- Scalarisation : AUC
→ Vision globale

	Predicted positive	Predicted negative	
Positive examples	TP	FN	Pos
Negative examples	FP	TN	Neg
	PPos	PNeg	N

Alternative proposée

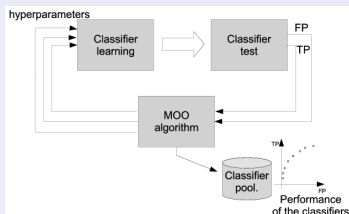
- Front ROC
 - ▶ Population de classifieurs
 - ▶ Vision locale



- Un problème d'optimisation multiobjectif

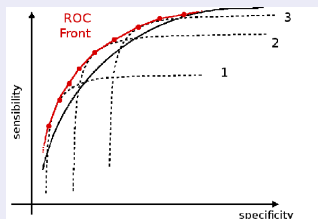
Optimisation multi-objectif : application (26)

Approche choisie [PR'10]



- Classifieurs SVM
- Paramètres γ , C_+ , C_-
- Algorithme d'optimisation : NSGA-II ([DEB'00])

Résultats UCI

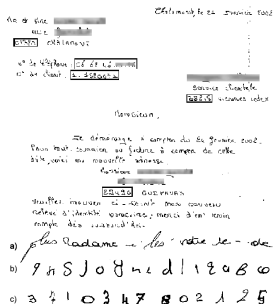


problem	Référence	AUC literature	AUF
australian	[Wu 05]	90.25 ± 0.6	96.22 ± 1.7
wdbc	[Ferri 02]	94.7 ± 4.6	99.59 ± 0.4
breast cancer	[Bostrom 05]	99.13	99.78 ± 0.2
ionosphere	[Rakoto 04]	98.7 ± 3.3	99.00 ± 1.4
heart	[Wu 05]	92.60 ± 0.7	94.74 ± 1.9
pima	[Cortes 04]	84.80 ± 6.5	87.42 ± 1.2

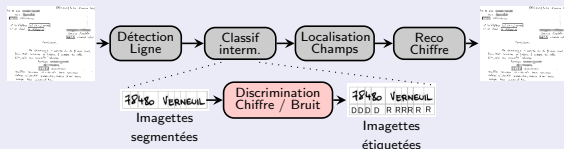
Optimisation multi-objectif : application (27)

Application sur un problème réel [Chatelain]

- Détection des champs numériques dans des courriers entrants



Le système global

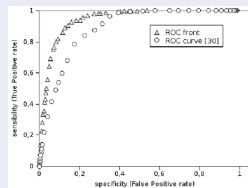
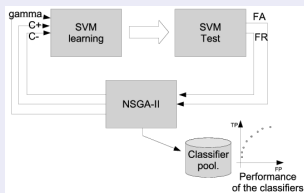


Coûts :

- déséquilibrés
- difficiles à estimer

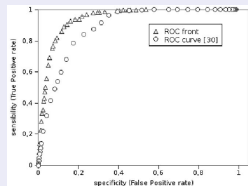
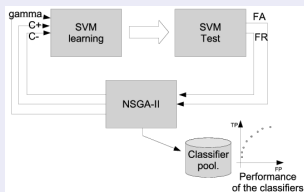
Optimisation multi-objectif : application (28)

Résultats

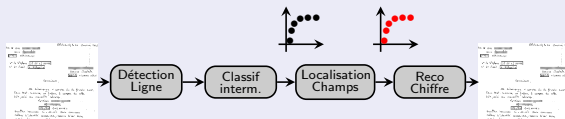


Optimisation multi-objectif : application (28)

Résultats



Bilan



TP rate	98.8	99.04	99.26	99.48	99.76	99.96	100
F1-Measure	0.17	0.19	0.22	0.25	0.32	0.30	0.23

- Population → évolutivité → changement dynamique de classifieur