Modèles Simplifiés et Apprentissage Actif

PhD Philippe Rolet

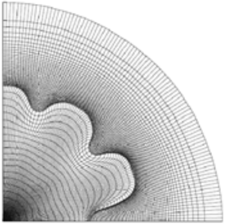
Michele Sebag
CNRS – INRIA – UPS

Jean-Marc Martinez
DMSS – CEA

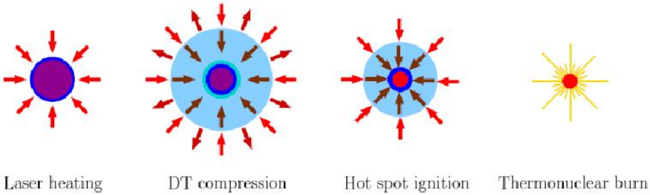
22 septembre 2008

Capitalisation du savoir-faire scientifique et industriel

Numerical Engineering



- De gros codes de calcul
- Chers en temps calcul
- Chers en expertise



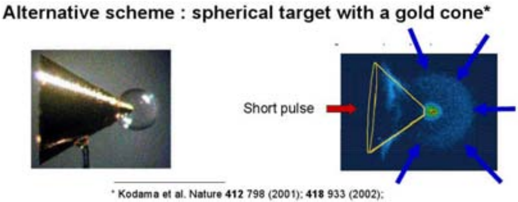
Fusion par confinement inertiel, ICF

Objectif: modèles simplifiés

Buts

- Un résultat approché
- Pour une fraction du temps de calcul
- Raccourcir le cycle de conception
- Conception optimale

More is Different



Apprentissage Statistique

Contexte

$$\text{Monde} \rightarrow \text{instance } \mathbf{x}_i \rightarrow \begin{array}{c}\text{Oracle}\\ \downarrow \\ y_i\end{array}$$

Apprentissage Supervisé

Input: Base d'apprentissage $\mathcal{E} = \{(\mathbf{x_i}, y_i), i = 1 \ldots n\}$

Output: Hypothèse h: $\mathbf{x} \mapsto h(x)$
Confiance: Estimation $E[h(x) \neq y]$

Critère: Minimiser $E[\ell(h(x) \neq y)]$ *espérance du coût de l'erreur*

ML et Programmation Adaptative

Quand

- Quand les spécifications sont inconnues *vision*
- Quand l'environnement est incertain *filtrage collaboratif*
- Dans un environnement ouvert *Mars*

Programmer: de moins en moins

1. Procédural
2. Déclaratif
3. Inductif
4. Guidé par les préférences
   Brain Computer Interface

Apprentissage Statistique

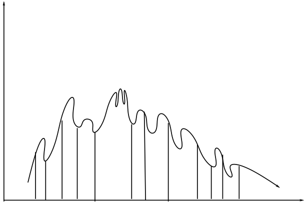
Minimiser l'espérance du coût de l'erreur

$$\text{Minimize } E[\ell(h(x), y)]$$

Principle

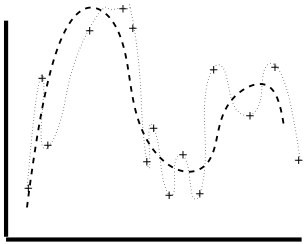
If h "does well" empirically, and h is "regular", h will do well in expectation.

$$E[F] \leq \frac{\sum_{i=1}^{N} F(x_i)}{n} + c(F, n)$$

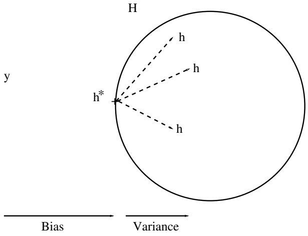


Apprentissage Statistique

Fitter les données



Biais variance



Machine Learning: current challenges

ML and Optimization

- Convex criteria
- Large scale optimization — Wshop ICML 2008

ML and Settings

- Unsupervised ML — *Autonomic Computing, EGEE-III*
- Semi-supervised
- Reinforcement learning — *MoGo*
- Active learning
- Learning to rank
- Multi-task learning
- Interactive learning — *DigiBrain*
- Distributed learning — *SYMBRION IP*

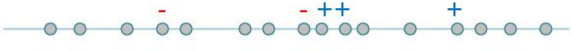
Active Machine Learning

Selecting examples

- Learning convergence: error decreases as $\sqrt{\frac{1}{N}}$ or $\frac{1}{N}$ for iid examples
- But not all examples are equally informative
- Find those enforcing a fast learning rate

Principle

- Design of Experiments
- Maximize variance — Angluin 89, Cohn et al. 94
- Maximize reduction of search space — Dasgupta 05



Active Machine Learning, 2

Aspects

- Separable or Agnostic ?
- Which distance among hypotheses ? L_∞, L_1, L_2

Broad cases

- H and covering number: if too large, quasi-random
- Else: Strategy to divide the space.
  - Splitting index *Dasgupta 2005*
  - Disagreement coefficient *Balcan et al. 2006* *Hanneke 2008*
- On-going:
  - Extension of *Dasgupta* to regression
  - Coupling with dimensionality reduction
  - Coupling with Multi-Armed Bandit

ML, Applications récentes

- Web and intranet information retrieval
- Spam detection
- Personalized medicine
- Compiler optimization
- Collaborative filtering
- Probabilistic robotics
- Brain Computer Interface, spelling, prothesis,
- ...

NoE PASCAL (2003-2008, 2008-2013)
Pattern Analysis, Statistical Modelling and Computational Learning