

Présentation de l'ES APPRENTISSAGE ARTIFICIEL (« Machine-Learning »)

Fabien Moutarde
Centre de Robotique (CAOR)
MINES ParisTech (ENSMP)

Fabien.Moutarde@mines-paristech.fr
<http://www.mines-paristech.fr/~moutarde>

Objet du cours et intervenants

- **Présenter :**
 - un panorama *assez large* des algorithmes récents dits d'apprentissage artificiel/statistique (réseaux neuronaux, SVM, dopage [boosting], méthodes probabilistes, k-means, Kohonen, AG,...) permettant des analyses et modélisations de données plus puissantes que les simples méthodes statistiques linéaires classiques
 - cadre théorique et méthodologique commun
 - divers types (classification, régression, clustering, ...) et domaines d'applications
- **Faire pratiquer en TP plusieurs de ces techniques**
- **Intervenants**
 - Fabien Moutarde (CAOR/MinesParis)
 - Jean-Philippe Vert (CBIO/MinesParis) + Toby Hocking (CBIO+INRIA)
 - Guillaume Obozinski (INRIA/ENS)

Programme prévisionnel

- **Séance 1 (vendredi 25 mars 2011 après-midi) : Intro et généralités sur l'apprentissage statistique + Réseaux de neurones à couches**
- **Séance 2 (vendredi 1 avril 2011 après-midi) : Apprentissage en Bio-Informatique, SVMs et méthodes à noyaux pour données structurées (graphes, etc...) : Jean-Philippe Vert + Toby Hocking (CBIO)**
- **Séance 3 (vendredi 08 avril 2011 après-midi) : Autres algorithmes d'apprentissage supervisé : Arbres de décision (ADB), dopage (boosting)**
- **Séance 4 (vendredi 29 avril 2011 après-midi) : Méthodes probabilistes (+ parcimonieuses ?) : Guillaume Obozinski (INRIA/ENS)**
- **Séance 5 (vendredi 13 mai 2011 *matin*) : Rappels + Apprentissage non supervisé, clustering + ACI**
- **Séance 6 (vendredi 20 mai 2010 *matin*) : Méthodes heuristiques (Algos Génétiques, Essaims Particulaires, ACO, ...)**

- **EXAM (vendredi 20 mai matin 10h-12h30) : Présentation orale du travail de mini-projet**

Evaluation : micro-projet

- **Micro-projet par groupe de 2 élèves**
(hors emploi du temps)

Principe proposé :

- choix d'un jeu de données parmi benchmarks publics
(pour lequel il existe au moins une publi avec perf. obtenues)
→ voir par ex. dans « UCI Machine-Learning repository »
- application + comparaison de diverses techniques d'apprentissage (eg celles pratiquées en TP)
- Rapport présentant et analysant les résultats,
+ comparaison à performances publiées
- Présentation orale (~12mn / groupe)

- *Apprentissage artificiel : concepts et algorithmes*, A. Cornuéjols, L. Miclet & Y. Kodratoff, Eyrolles, 2002.
- *Pattern recognition and Machine-Learning*, Christopher M. Bishop, Springer, 2006.
- *Introduction to Data Mining*, P.N. Tan, M. Steinbach & V. Kumar, AddisonWesley, 2006.
- *Machine Learning*, Thomas Mitchell, McGraw-Hill Science/Engineering/Math, 1997.

1. **Intro sur l'apprentissage statistique supervisé (copies de transparents FM)**
2. **Réseaux de neurones à couches (copies de transparents FM)**
3. **Introduction aux Arbres de Décision (copies de transparents FM)**
4. **Principe du boosting, et application à la reconnaissance visuelle de catégorie d'objet avec l'outil SEVILLE (copies de transparents FM)**
5. **Brève introduction aux SVMs (copies de transparents FM)**
6. **Support Vector Machines, kernels, and applications in computational biology (copies de transparents de Jean-Philippe Vert du CBIO/MINES ParisTech)**
7. **Méthodes probabilistes (copies de transparents de Guillaume Obozinski, INRIA/ENS)**
8. **Formalisme apprentissage supervisé + intro apprentissage NON-supervisé (copies de transparents FM)**
9. **Algorithmes génétiques et autres meta-heuristiques (copies de transparents FM)**