-1

Théorie de l'apprentissage

Session organisée par Nicolas Vayatis

Le domaine de la théorie statistique de l'apprentissage a été le témoin d'avancées significatives ces dernières années. En effet, la nature des contributions théoriques a mis en lumière des applications originales des techniques de processus empiriques et des inégalités de concentration et ont définitivement ancré la recherche en apprentissage dans le champ de la statistique mathématique. Ces contributions ont essentiellement exploré, pour l'instant, des modèles et des critères de performances qui ne prennent pas en compte les problématiques des praticiens de l'apprentissage computationnel et de l'analyse de données. Le nouveau défi pour la théorie de l'apprentissage est de s'ouvrir aux questions plus concrètes d'interprétabilité des modèles, de calibrage des procédures statistiques proposées, d'élaboration de mesures de performance pratiques. Le but de cette session est précisément d'illustrer ces nouvelles directions de recherche à travers les exposés de jeunes chercheurs en apprentissage statistique.

Références :

[1] *Theory of Classification : a Survey of Recent Advances,* (2005) S. Boucheron, O. Bousquet, and G. Lugosi, ESAIM : Probability and Statistics.

Adresse de l'organisateur :

Nicolas Vayatis
Laboratoire de Probabilités et Modèles Aléatoires
Université Paris 6
Case courrier 188
4, place Jussieu
75252 Paris Cedex 05, France
E-mail : `vayatis@ccr.jussieu.fr`
<`http://www.proba.jussieu.fr/pageperso/vayatis/main.html`>

Some theoretical and practical problems

par **Olivier Bousquet**

The idea of this talk is to investigate two totally opposed directions : the theoretical foundations of learning on one side, and the practical use of learning algorithms on the other side. We will try to make a list of theoretical questions that may be relevant for the understanding and design of algorithms, in particular focusing on the case where almost no assumptions on the way the data is generated are needed (if any). We will discuss the relevance of this setting in practical problems, and then move to a discussion about applications of learning algorithms to solving engineering problems. We will explain which criteria might be relevant and show on examples that the answer is sometimes very far from what existing algorithms address.

Adresse :

Olivier Bousquet
Pertinence
32, rue des Jeûneurs
75002 Paris, France
E-mail : `olivier.bousquet@pertinence.com`
<`http://www.kyb.mpg.de/~bousquet`>

Sélection itérative de caractéristiques en régression

par **Pierre Alquier**

On s'intéresse ici à l'estimation de la régression : on essaie de prévoir une variable Y comme fonction d'une variable X, X étant représentée par un ensemble de m caractéristiques ($f_1(X), ..., f_m(X)$, les fonctions $f_1, ..., f_m$ étant des ondelettes, des fonctions définies à l'aide de noyaux...). On dispose pour cela d'un échantillon de taille N de couples (X, Y), avec possiblement $N < m$. Dans un premier temps, on donne une borne sur le risque d'un estimateur dans chacun des m modèles linéaires définis en utilisant une seule caractéristique. Cette borne motive une méthode itérative de sélection et d'agrégation de caractéristiques pour laquelle on peut prouver que chaque caractéristique ajoutée améliore effectivement les performances de l'estimateur obtenu. Les résultats sont présentés dans un premier temps dans le cadre "transductif", c'est-à-dire où l'on essaie d'évaluer la fonction de régression en un ensemble fini de points d'intérêts. On présente ensuite les résultats dans le cadre inductif, qui nécessite ici plus d'hypothèses, sous lequelles on démontre que l'estimateur obtenu est adaptatif par rapport à la régularité de la fonction.

Adresse :

Pierre Alquier
CREST
-ENSAE
Timbre J340
3, avenue Pierre Larousse
92240 Malakoff, France
E-mail : `Pierre.Alquier@ensae.fr`
<`http://www.crest.fr/pageperso/alquier/alquier.htm`>

Le problème du ranking

par **Stéphan Clémençon**

Dans les applications liées à la fouille de documents, le but n'est pas seulement de construire un classifieur décidant si un document x dans une liste $\mathcal{X}$ est pertinent ou non, mais d'apprendre une fonction de score $s : \mathcal{X} \to \mathbb{R}$ permettant d'ordonner tous les documents par rapport à leur pertinence. Le problème du ranking bipartite se réduit à la classification binaire de données dépendantes où la performance est mesurée grâce au critère AUC. L'estimateur naturel du risque est de la forme d'une U-statistique et on dérive la consistance des méthodes basées sur la minimisation du risque empirique en utilisant la théorie des U-processus. Dans ce contexte, il reste possible d'obtenir des vitesses de convergence rapides sous des hypothèses de bruit générales.

Adresse :

Stéphan Clémençon
MODAL'X
Université Paris 10
200, avenue de la République
92000 Nanterre, France
E-mail : sclemenc@u-paris10.fr

Descente miroir, agrégation et sélection de modèle

par **Philippe Rigollet**

La descente miroir est une méthode d'optimisation stochastique permettant d'agréger des fonctions (estimateurs ou classifieurs). Etant donnée une collection de fonctions, l'agrégation permet d'imiter la meilleure combinaison convexe de ces fonctions (agrégation convexe) ou la meilleure de ces fonctions (agrégation par sélection de modèle) pour une perte donnée. Ceci se traduit par des inégalités d'oracle exactes qui ont pour conséquence directe la construction d'estimateurs non paramétriques adaptatifs (à la constante près) pour différents problèmes tels que la régression, la classification et l'estimation de densité.

Adresse :

Philippe Rigollet
Laboratoire de Probabilités et Modèles Aléatoires
Université Paris 6
Case courrier 188
4, place Jussieu
75252 Paris Cedex 05, France
E-mail : rigollet@ccr.jussieu.fr

Performances statistiques d'algorithmes d'apprentissage : "Kernel Projection Machine" et analyse en composantes principales à noyau

par **Laurent Zwald**

Pendant cet exposé, des méthodes statistiques seront utilisées pour résoudre des problèmes d'apprentissage. Dans un premier temps, l'analyse en composantes principales à noyau (KPCA) est etudiée statistiquement. Il s'agit d'un algorithme de réduction de la dimension tres utilisé en pratique. Puis, un nouvel algorithme de classification est proposé : la Kernel Projection Machine (KPM). Il est etudié de façon théorique et ses performances ont été testées sur des données. Tout en s'inspirant des "Supports Vectors Machine" (SVM), il utilise un procédé de régularisation fini-dimensionnelle. On verra comment il est relié a la KPCA.

Adresse :

Laurent Zwald
Laboratoire de Probabilités et Statistiques
Université Paris 11 Bâtiment 425
91405 Orsay Cedex, France
E-mail : laurent.zwald@math.u-psud.fr
<http://www.math.u-psud.fr/~zwald/>