

## ***Société organisatrice***

La SFdS organise depuis quatre ans maintenant des cours thématiques annuels destinés au monde professionnel : *les Ateliers Statistiques de la SFdS*. Leur principal objectif est de faire profiter les statisticiens et les utilisateurs de la statistique de démarches et de méthodes qui leur permettront d'améliorer leurs connaissances et, de fait, leur capacité à résoudre des problèmes posés dans leur entreprise.

Le thème de cette année est la théorie de l'apprentissage et ses applications.

Le premier cours organisé sur ce thème est une introduction aux bases théoriques de la modélisation prédictive, aux méthodes neuronales et aux SVM (Support Vector Machines). La théorie statistique de l'apprentissage est née de travaux à la frontière entre la statistique et les sciences de l'information (Intelligence artificielle, « machine learning ») et propose des algorithmes et des concepts nouveaux. La théorie de l'apprentissage fournit des bornes pour l'erreur de prédiction en même temps qu'une stratégie de constructions de modèles dont la complexité peut croître avec le nombre de données. La notion de modèle issu des données diffère ici de l'acception usuelle et le choix entre modèles se fait sur leurs capacités prédictives. Les techniques de réseaux de neurones, largement développées à partir des années 80, ont permis d'obtenir rapidement des résultats sur des problèmes de data mining significatifs. Depuis leur apparition, les algorithmes SVM ont été déclinés dans divers contextes (régression, détection d'anomalies, analyse des données) et, très récemment, leurs performances statistiques ont été comprises grâce à l'étude du principe de minimisation de risques convexifiés.

Au cours de la première journée, les bases de la théorie issue des travaux de V. Vapnik et ses liens avec des aspects plus classiques seront fournies. Puis les techniques de réseaux de neurones seront présentées. Il sera notamment montré, d'une part, comment de nombreuses techniques de régularisation mises en œuvre pour rendre efficaces les réseaux de

neurones peuvent être replacées dans le cadre théorique de la SRM (Structural Risk Minimization) de Vapnik et, d'autre part, comment l'industrialisation systématique de la SRM permet de mettre en œuvre une nouvelle approche du data mining applicable à des problèmes de très grande taille (milliers de variables et millions d'enregistrements). Des exemples du marketing en utilisant le logiciel KXEN permettront d'illustrer ces techniques.

La seconde journée approfondira les SVM. Notamment, l'ensemble des aspects théoriques et pratiques des méthodes à noyaux en apprentissage sera décrit et illustré sur des exemples issus du domaine biomédical et de la reconnaissance de formes.

Ce cours sera animé par **Gilbert Saporta** (CNAM), **Françoise Soulié-Fogelman** (KXEN) et **Nicolas Vayatis** (Université Paris 6).

### **les 22 et 23 juin 2006**

**9h00-12h30 et 14h00-17h30**

Afin d'assurer une plus grande convivialité à la formation et de favoriser les échanges entre les participants, **le nombre de places est limité à 30.**

Date limite d'inscription : **9 juin 2006**

### ***Plan de la formation***

#### **Jeudi 22 juin**

##### **Introduction générale**

- Positionnement scientifique et historique : machine learning et statistique
- Panorama des méthodes d'apprentissage supervisé et non supervisé
- Le dilemme biais-variance
- VC dimension et SRM
- Régularisation et généralisation
- Autres critères de choix de modèles (globaux et locaux)
- Approche bayésienne

#### **Réseaux de neurones**

- Présentation des différentes méthodes de réseaux de neurones
- Réseaux de neurones et statistiques
- Limitations pratiques des réseaux de neurones
- Présentation de différentes techniques de régularisation pour rendre efficaces les réseaux de neurones
- Application de la théorie SRM (Structural Risk Minimization) de Vapnik aux réseaux de neurones
- Industrialisation de la méthode SRM dans le cas de régressions polynomiales
- Un exemple : utilisation du logiciel KXEN

#### **Vendredi 23 juin**

- Le problème statistique de la classification binaire
- Algorithme de séparation linéaire à marge optimale
- Rappels d'optimisation convexe
- Astuce du noyau et algorithme SVM
- Présentation de logiciels
- Exemples de classification pour des données artificielles
- Risques convexifiés et analyse des performances des SVM
- SVM pour la régression et la détection d'anomalies
- Noyaux définis positifs et distances associées
- Méthodes à noyaux pour l'analyse des données : le cas de l'ACP à noyau
- Applications sur des données réelles

### ***Organisation***

Le cours se déroulera à l'IHP, 11 rue Pierre et Marie Curie 75005 Paris.  
Accès : RER ligne B station : Luxembourg  
Bus : 21, 27, 38, 84, 85, 89

A retourner au secrétariat de la SFdS, IHP, 11 rue Pierre et Marie Curie, 75231 Paris cedex 05.  
**A noter que l'inscription ne sera définitive qu'à réception du règlement ou d'un bon de commande.**

NOM.....Prénom.....  
Fonction.....Organisme.....  
Adresse complète.....  
.....  
Code Postal.....Ville :.....  
Pays.....Tél.....Fax.....E-mail.....

souhaite m'inscrire au cours : Introduction à l'apprentissage statistique. Méthodes et applications.  
organisé les **22 et 23 juin 2006** à l'IHP

o verse     o m'engage à verser     o par chèque     o par virement     o 550 €

DATE et SIGNATURE :

## **Droits d'inscription au cours**

**550 €\* pour les membres SFdS**

\*Pour des raisons juridiques, les cours de la SFdS doivent s'adresser uniquement à ses membres. Les non membres de l'association doivent s'acquitter de leur adhésion pour l'année en cours pour pouvoir participer aux formations. Merci dans ce cas de prendre contact avec le secrétariat.

### **Les droits d'inscription au cours comprennent :**

- la participation à la formation
- les documents de cours
- les pauses café
- les repas du midi.

### **Annulation**

Le remboursement des droits d'inscription sera possible sur demande écrite transmise avant le 9 juin 2006 et donnera lieu à une retenue de 150 € pour frais. Aucun remboursement ne sera effectué après cette date.

### **Modalités de règlement**

Le règlement des droits d'inscription à la formation peut se faire par :

- **chèque bancaire** ou **chèque postal** émis sur une banque française, libellé à l'ordre de la **Société Française de Statistique**
- **virement bancaire**
- **bon de commande**

### **Numéro d'agrément Formation Continue SFdS :**

117 531 085 75

La SFdS n'est pas assujettie à TVA



## **Atelier Statistique**

***"Introduction à  
l'apprentissage statistique.  
Méthodes et applications"***

**PARIS (I.H.P.)**

**les 22 et 23 juin 2006**

Tél : (33) 01 44 27 66 60  
Fax : (33) 01 44 07 04 74

e-mail : [sfds@ihp.jussieu.fr](mailto:sfds@ihp.jussieu.fr)  
Site internet : <http://www.sfds.asso.fr>