



Université
de Lomé



CIMPA

RESUME DES COURS

ECOLE DE RECHERCHE CIMPA

Thème :

**Approches d'apprentissage automatique –
Application aux données de santé**

Date et lieu : 12 au 23 juillet 2021 à l'Université de Lomé (Togo)

Table des matières

Cours 1 : Mathématiques pour l'apprentissage statistique	3
Cours 2 : Apprentissage statistique : théorie et algorithmes.....	4
Cours 3 : Machine Learning avancé : contrôle de la complexité et adaptation	5
Cours 4 : Intelligence artificielle : apprentissage profond	6
Cours 5 : Algorithmique Avancée	7
Cours 6 : Machine Learning pour la santé	8
Cours 7 : Systèmes et techniques intelligentes de collecte de données	9

Cours 1 : Mathématiques pour l'apprentissage statistique

Durée : 8 heures

Intervenants :

- CANU Stéphane (Professeur), INSA Rouen Normandie (France)
- GNEYOU Kossi (Professeur), Université de Lomé (Togo)

Résumé du cours

Le cours propose de (re)voir les outils mathématiques pour l'apprentissage statistique. Dans sa première partie, ce cours introduira les fondements des statistiques et la théorie des probabilités. Il s'attachera dans la deuxième partie à présenter les problèmes d'optimisation, les algorithmes de résolution et l'analyse de leurs propriétés. Le cours mettra l'accent sur l'articulation de ces outils mathématiques avec les méthodes et les fondements théoriques de l'apprentissage statistique.

Les notions suivantes seront approfondies :

- Statistiques et probabilités
- Optimisation avec et sans contraintes, convexe et non-convexe
- Méthodes de descentes, méthodes proximales
- Analyse des propriétés de convergence

Des séances d'exercices et de travaux pratiques en Python ou R compléteront le cours.

Cours 2 : Apprentissage statistique : théorie et algorithmes

Durée : 8 heures

Intervenants :

- Liva RALAIVOLA (Professeur), Institut Universitaire de France - Criteo AI Lab (France)
- N'GUESSAN Assi (Maître de Conférences, HDR), Université de Lille (France)

Résumé du cours

L'apprentissage automatique, Machine Learning, est l'objet d'une attention phénoménale aussi bien de la part de la communauté de la recherche universitaire que de la part de grands acteurs industriels (Google, Facebook, Amazon, Apple, Microsoft...) qui la mettent au cœur de leurs actions en recherche et développement pour proposer des systèmes d'intelligence artificielle de plus en plus performants.

Les fondements théoriques de cette discipline se trouvent dans les travaux de Vapnik, à la fin des années 70, qui venaient notamment caractériser les propriétés dites de généralisation de modèles appris à partir d'échantillons de données finis. Ces fondements reposent sur des outils de statistique non paramétrique, de combinatoire et d'algorithmique (pour l'optimisation, en particulier). Depuis ces premiers travaux, de nombreuses avancées sont venues enrichir la théorie de l'apprentissage et ont donné lieu au développement de méthodes et algorithmes d'apprentissage automatique.

L'objet de ce cours est de visiter certaines de ces avancées tant algorithmiques que théoriques. Nous discuterons les principaux modèles d'apprentissage statistique verrons et comment des éléments de théorie s'articulent avec ces modèles. Les notions suivantes seront abordées :

- Panorama des problèmes d'apprentissage statistique
- Méthodes à noyaux
- Méthodes ensemblistes (Bagging, Boosting, Gradient boosted trees, Random forest ...)
- Algorithmes de bandits
- Théorie de l'apprentissage statistique : inégalités de concentration, bornes de généralisation

Des exercices et des travaux pratiques (en Python) compléteront ce cours.

Cours 3 : Machine Learning avancé : contrôle de la complexité et adaptation

Durée : 9 heures

Intervenants

- Alain RAKOTOMAMONJY (Professeur), Université de Rouen - Criteo AI Lab (France)
- Gilles GASSO (Professeur), INSA Rouen Normandie (France)

Résumé du cours :

Dans de nombreux domaines d'application tels que la biologie computationnelle, le text mining, les données à partir desquelles les modèles de prédiction sont appris, sont systématiquement de grande dimension. Une façon de traiter de tels types de données consiste à apprendre des modèles dont la complexité est contrôlée à travers la régularisation parcimonieuse. Les modèles parcimonieux permettent non seulement de reconstruire des signaux de grande dimension, de découvrir les variables pertinentes mais aussi d'autoriser une interprétation plus aisée des modèles appris. La parcimonie rend également l'apprentissage des modèles plus efficace, conduisant à des solutions efficaces à partir d'un nombre limité de données.

Un autre aspect important des modèles prédictifs est leur capacité d'adaptation à de nouvelles données lorsque l'hypothèse i.i.d. n'est plus valide. Des approches récentes d'adaptation de modèles ont récemment émergé en Machine Learning en se basant sur la théorie du Transport Optimal.

Ce cours couvrira la théorie, les algorithmes et les applications de la parcimonie des modèles et leur adaptation à des domaines d'application connexes. Le programme inclura :

- Apprentissage statistique régularisé
- Optimisation non lisse, algorithme de gradient proximal
- Application à l'apprentissage de modèles interprétables
- Transport Optimal, problème d'affectation, régularisation entropique
- Application à l'adaptation de domaine, transfer learning

Des exercices théoriques et pratiques (en Python) compléteront le cours.

Cours 4 : Intelligence artificielle : apprentissage profond

Durée : 10 heures

Intervenants :

- **Morgane Rivière** (Researcher), Facebook AI Research (FAIR) (France)
- **Olivier TEYTAUD** (Senior Researcher), Facebook AI Research (FAIR) (France)

Résumé du cours

On présentera les grandes lignes de l'apprentissage profond et leurs applications :

- Réseaux convolutionnels,
- Réseaux récurrents,
- Réseaux résiduels,
- Architectures "légères" (pour infrastructures légères de calcul)
- Apprentissage des réseaux : Relu, Régularisation, Dropout, optimisation sans dérivée, méthodes stochastiques d'optimisation
- Optimisation d'hyper-paramètres, construction automatique de réseaux profonds,
- Théorie de généralisation des architectures profondes
- Application : computer vision, traitement du langage naturel, apprentissage par renforcement

Des travaux pratiques (en python) basés sur les libraires Keras ou Pytorch compléteront ce cours.

Cours 5 : Algorithmique Avancée

Durée : 4 heures

Intervenant : **RAVELOMANANA Vlady** (Professeur), IRIF – UMR8243 – Université Denis Diderot Paris7 (France)

Résumé du cours

Dans un premier temps, ce cours propose de voir plusieurs concepts algorithmiques avancés comme l'exploration exhaustive, le backtracking, les algorithmes gloutons et la programmation dynamique.

Ensuite, on étudiera les algorithmes fondamentaux issus de problèmes de l'informatique fondamentale classiques mais aussi ceux intimement liés avec les autres cours de l'Ecole :

- Algorithmes de graphes (coloration, clique, stable, coupe, ...)
- Algorithmes randomisés (Monte-Carlo, Las Vegas, analyse en moyenne)
- Algorithmes distribués dans les systèmes complexes (collecte de données, diffusion, ...)
- Algorithmes d'approximation (schéma d'approximation, inapproximabilité)
- Algorithmes online

Des exercices compléteront ce cours.

Cours 6 : Machine Learning pour la santé

Durée : 10 heures

Intervenants :

- **Chloé-Agathe AZENCOTT** (Professeur), Centre for Computational Biology (CBIO) of Mines ParisTech, Institut Curie and INSERM (France)
- **Fifonsi Diane GBEASSOR-KOMLANVI**, Département de Santé Publique, Université de Lomé (Togo)

Résumé du cours

Ce cours se veut à l'interface de l'apprentissage automatique et de la biologie-santé. Il introduira aux spécificités des données de santé, les problématiques de traitement statistique qu'elles posent et les approches de solution par l'apprentissage automatique. Les thèmes suivants seront abordés :

- Données de santé : un regard croisé Nord-Sud sur les enjeux et les contraintes pratiques,
- Complexité des données médicales et stratégies de modélisation de la connaissance
- Modélisation d'interactions biologiques, modélisation de petites molécules par les méthodes à noyaux
- Applications génomiques : sélection de variables en grande dimension, régularisation, tests statistiques de consistance
- Biologie computationnelle et graphes : apprentissage statistique sur graphes (méthodes à noyaux, deep learning), applications à la chemoinformatique et aux réseaux biologiques.

Cours 7 : Systèmes et techniques intelligentes de collecte de données

Durée : 4 heures

Intervenant : **ADJALLAH Kondo**, Université de Lorraine (France)

Résumé du cours :

Ce cours vise à donner un ensemble de méthodes et techniques de modélisation, de fiabilisation et d'optimisation d'un processus de collecte de données intervenant en amont de l'exploitation des données. Les éléments du contenu aideront à assurer les conditions de qualité des modèles qui en seront extraits. Le cours sera structuré en 3 parties focalisées respectivement sur le processus de collecte des données, sa fiabilisation et sa validation, avec des applications orienté santé. Les interventions s'appuieront, entre autres, sur des méthodes d'apprentissage et d'optimisation de type biomimétique ainsi que sur la théorie des ensembles flous.

- Systèmes et processus de collecte des données
- Modèles de réseaux d'agents de collecte de données
- Modélisation et Fusion de données
- Méthodes de fiabilisation et optimisation du processus de collecte des données
- Modèles d'optimisation du processus de collecte par apprentissage floue
- Apprentissage automatique et intégration de données de santé
- Algorithmes avancés de classification
- Approche biomimétique d'optimisation des classifications
- Construction de modèle graphique à partir de données
- Application en travaux pratiques
- Modèles d'évaluation de la fiabilité d'un processus de collecte
- Modèles adaptatifs d'apprentissage de la collecte et modèle de complexité
- Validation des processus et modèles collecte des données
- Application en travaux pratiques