COURSES SUMMARY

CIMPA RESEARCH SCHOOL

Theme in french:

Approches d'apprentissage automatique – Application aux données de santé

Theme in english:

Machine learning approaches – Application to health data

Dates and location: **12 to 23 july 2021** at **Université de Lomé (Togo)**

Contents

Course 1: Mathematics for Machine learning

Duration: 8 hours

Teachers:

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

Summary

The course offers to see (or revisit) the mathematical tools for machine learning. In the first part, the foundations of statistics and probability theory will be introduced. In the second part, the course will focus on presenting the optimization problems, the resolution algorithms and the analysis of their properties. The course will focus on the articulation of these mathematical tools with the methods and theoretical foundations of machine learning.

The following notions will be studied in depth:

- Statistics and probability
- Optimization with and without constraints, convex and non-convex
- Descent methods, proximal methods
- Analysis of convergence properties

Exercises and practical work in Python or R will complete the course.

Course 2: Machine learning: theory and algorithms

Duration: 8 hours

Teachers:

- **Liva RALAIVOLA** (Professor), Institut Universitaire de France - Criteo AI Lab (France)
- **N'GUESSAN Assi** (Associate Professor), Université de Lille (France)

Summary

Machine learning is the subject of phenomenal attention both from the university research community and from major industrial groups (Google, Facebook, Amazon, Apple, Microsoft, ...) who put it at the heart of their research and development actions to offer increasingly efficient artificial intelligence systems.

The theoretical foundations of this discipline can be found in the work of Vapnik at the end of the 1970s, which notably characterized the so-called generalization properties of models learned from finite data samples. These foundations are based on tools of nonparametric statistics, combinatorics and algorithms (for optimization, in particular). Since this early work, many advances have enriched learning theory and have given rise to the development of machine learning methods and algorithms.

The object of this course is to visit some of these advances, both algorithmic and theoretical. We will discuss the main models of machine learning and see how elements of theory relate to these models. The following concepts will be discussed:

- Overview of machine learning problems
- Core methods
- Set methods (Bagging, Boosting, Gradient boosted trees, Random forest ...)
- Bandit algorithms
- Theory of machine learning: concentration inequalities, generalization bounds

Exercises and practical work (in Python) will complete this course.

Course 3: Advanced Machine Learning: complexity control and adaptation

Duration: 9 hours

Teachers

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

Summary:

In many fields of application such as computational biology, text mining, the data from which prediction models are learned are systematically large. One way to deal with such types of data is to learn models whose complexity is controlled through parsimonious regularization. Parsimonious models not only make it possible to reconstruct large-scale signals, to discover the relevant variables but also to allow an easier interpretation of the learned models. Parsimony also makes learning models more efficient, leading to efficient solutions from a limited amount of data.

Another important aspect of predictive models is their ability to adapt to new data when the i.i.d. assumption is no longer valid. Recent model adaptation approaches have recently emerged in Machine Learning based on Optimal Transport theory.

This course will cover the theory, algorithms and applications of model parsimony and their adaptation to related application areas. The program will include:

- Regularized machine learning
- Non-smooth optimization, proximal gradient algorithm
- Application to the learning of interpretable models
- Optimal transport, assignment problems, entropy regularization
- Application to domain adaptation, transfer learning

Exercises and practical work (in Python) will complete this course.

Course 4: Artificial Intelligence: deep learning

Duration: 10 hours

Teachers:

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

Summary

We will present the main lines of deep learning and their applications:

- Convolutional networks,
- Recurring networks,
- Residual networks,
- “Light” architectures (for light computing infrastructures)
- Network learning: Proofread, Regularization, Dropout, derivative-free optimization, stochastic optimization methods
- Optimization of hyper-parameters, automatic construction of deep networks,
- Generalization theory of deep architectures
- Application: computer vision, natural language processing, reinforcement learning

Practical work (in Python) based on the Keras and Pytorch packages will complete this course.

Course 5: Advanced Algorithmics

Duration: 4 hours

Teacher: **RAVELOMANANA Vlady** (Professor), IRIF – UMR8243 – Université Denis Diderot Paris7 (France)

Summary

As a first step, this course proposes to see several advanced algorithmic concepts such as exhaustive exploration, backtracking, greedy algorithms and dynamic programming.

Then, we will study the fundamental algorithms resulting from classic fundamental computer science problems but also those closely linked with the other courses of the School:

- Graph algorithms (colouring, click, stable, cut, ...)
- Randomized algorithms (Monte-Carlo, Las Vegas, average analysis)
- Distributed algorithms in complex systems (data collection, dissemination, ...)
- Approximation algorithms (approximation scheme, inapproximability)
- Online algorithms

Exercises will complete this course.

Course 6: Machine Learning for health

Duration: 10 hours

Teachers:

- **Chloé-Agathe AZENCOTT** (Professor), Centre for Computational Biology (CBIO) of Mines ParisTech, Institut Curie and INSERM (France)
- **Fifonsi Diane GBEASSOR-KOMLANVI**, Department of Public Health, Université de Lomé (Togo)

Summary

This course is intended to be the interface of machine learning and health biology. It will introduce the specificities of health data, the statistical processing issues they pose and solution approaches through machine learning. The following topics will be discussed:

- Health data: a North-South perspective on the challenges and practical constraints,
- Complexity of medical data and knowledge modelling strategies
- Modelling of biological interactions, modelling of small molecules by core methods
- Genomic applications: selection of large-dimensional variables, regularization, statistical consistency tests
- Computational biology and graphs: statistical learning on graphs (kernel methods, deep learning), applications to chemoinformatics and biological networks.

Course 7: Intelligent systems and techniques for data collection

Duration: 4 hours

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

Summary:

This course aims to provide a set of modelling, reliability and optimization methods and techniques for a data collection process occurring upstream of data exploitation. The elements of the content will help to ensure the quality conditions of the models that will be extracted from it. The course will be structured in 3 parts focused respectively on the data collection process, its reliability and its validation, with health-oriented applications. Interventions will be based, among other things, on biomimetic-type learning and optimization methods as well as on fuzzy set theory.

- Data collection systems and processes
- Models of data collection agent networks
- Modelling and data fusion
- Methods for improving the reliability and optimization of the data collection process
- Models for optimizing the collection process by fuzzy learning
- Machine learning and health data integration
- Advanced clustering algorithms
- Biomimetic approach to optimization of clustering
- Construction of graphical model from data
- Application through practical work
- Models for assessing the reliability of a collection process
- Adaptive collection learning models and complexity model
- Validation of data collection processes and models
- Application trough practical work