DS3 DATA SCIENCE SUMMER SCHOOL JUNE 24-28 2019

PLENARY LECTURES

Optimization for large-scale machine learning

Francis BACH *- Inria -*

Schedule: June 25

Abstract: In this class I cover the main algorithms used in large-scale supervised learning problems such as logistic regression, with a particular focus on stochastic gradient methods and algorithms that come with convergence guarantees.

Advances and challenges in machine learning for personalised healthcare

Danielle BELGRAVE *- Microsoft Research, Cambridge, UK -*

Schedule: June 25

Abstract: Health presents one of the most challenging domains of machine learning and data science research. This offers an exciting opportunity for machine learning techniques to impact healthcare in a meaningful way. In this talk, we will look at some of the key challenges in machine learning for healthcare and some of the advances which have been made towards addressing these challenges.

AI and computational approaches: Implications for developmental psychopathology

David COHEN *- APHP - Hospital La Pitié Salpétrière -*

Schedule: June 25

Abstract: Leaving aside the exploration of large mental health databases of, AI and computational approaches are also a revolution in the field of developmental psychopathology. We will illustrate how the use of computational methods for the analysis and processing of social signals has allowed a paradigm shift to "see what cannot be seen". From various sources (home movies, laboratory experiments), we will show how these methods were used (1) to assess the interactive dynamics between partners (baby, child, parent, agent or robot) in a pathological context (autism, negligent mother) or not; (2) to characterize interpersonal exchanges and social signal turn tacking; (3) to introduce new approaches such as developmental robotics. We will end this presentation with a quick overview of the possible applications, especially in the field of specialized education and serious games.

DS3 DATA SCIENCE SUMMER SCHOOL JUNE 24-28 2019

Mathematical mysteries of deep neural networks

Stéphane MALLAT - *Collège de France* -

Schedule: June 24

Abstract: Deep neural networks obtain impressive results for image, sound and language recognition or to adress complex problems in physics. They are partly responsible of the renewal of artificial intelligence. Yet, we do not understand why they can work so well and why they fail sometimes, which raises many problems of robustness and explainability.
Recognizing or classifying data amounts to approximate phenomena which depend on a very large number of variables. The combinatorial explosion of possibilities makes it potentially impossible to solve. One can learn from data only if the problem is highly structured. Deep neural networks appear to take advantage of these unknown structures. Understanding this "architecture of complexity" involves many branches of mathematics and is related to open questions in physics. I will discuss some approaches and show applications.

Computational sociolinguistics

Rada MIHALCEA - *University of Michigan* -

Schedule: June 26

Abstract: Computational linguistics has come a long way, with many exciting achievements along several research directions, ranging from morphology and syntax to semantics and pragmatics. Simultaneously, there has been a tremendous growth in the amount of social media data available on web sites such as Blogger, Twitter, or Facebook, with all of these data streams being rich in explicit demographic information, such as the age, gender, industry, or location of the writer, as well as implicit personal dimensions such as personality and values. In this talk, I will describe recent research work undertaken in the Language and Information Technologies group at the University of Michigan, under the broad umbrella of computational sociolinguistics, where language processing is used to gain new insights into people's values, behaviors, and world views, and conversely information about people is leveraged to build better computational linguistic models. I will share the lessons learned along the way, and take a look into the future of this emerging research area.

Introduction to distributional reinforcement learning

Rémi MUNOS - *Google DeepMind* -

Schedule: June 24

Abstract: TBA

Computer vision and machine learning for environmental monitoring

Konrad SCHINDLER - *ETH Zürich* -

Schedule: June 25

Abstract: I will give on overview of our activities in the area of earth observation and environmental monitoring based on a wide range of images acquired from satellites, airplanes, mapping vehicles, and hand-held cameras. I will show different examples of how machine learning is applied to analyse imagery at large scale and extract environmentally relevant information, such as for example the locations, species and health of trees; the height and density of forests and agricultural plantations; the freezing and thawing of lakes; and the water levels during floods.

Action recognition

Cordelia SCHMID - *Inria* -

Schedule: June 25

Abstract: TBA

Low rank tensor methods in high dimensional data analysis

Ming YUAN - *Columbia University*-

Schedule: June 24

Abstract: Large amount of multidimensional data in the form of multilinear arrays, or tensors, arise routinely in modern applications from such diverse fields as chemometrics, genomics, physics, psychology, and signal processing among many others. At the moment, our ability to generate and acquire them has far outpaced our ability to effectively extract useful information from them. There is a clear demand to develop novel statistical methods, efficient computational algorithms, and fundamental mathematical theory to analyze and exploit information in these types of data. In this talk, I will review some of the recent progresses and discuss some of the present challenges.

Safe learning-based control

Melanie ZEILINGER - *ETH Zürich* -

Schedule: June 26

Abstract: A new opportunity for pushing the performance of emerging complex, large-scale and variable control systems to the next level is offered by the capability of learning from data during closed-loop operation. Safety concerns when integrating learning in a closed-loop, automated decision-making process, however, represent a key limitation for leveraging this potential in many industrial applications. This tutorial will discuss control techniques that ensure satisfaction of safety constraints while learning from data. In particular, we will discuss a framework to augment any learning-based controller with safety certificates. The ideas will be highlighted with application examples.