

**9e Journée Statistique et
Informatique pour la Science
des Données à Paris-Saclay**

**Rapport sur les
contributions**

ID de Contribution: 1

Type: **Non spécifié**

Reinforcement Learning, an Introduction and Some Results

mercredi 3 avril 2024 10:00 (50 minutes)

Reinforcement Learning is the “art” of learning how to act in an environment that is only observed through interactions.

In this talk, I will provide an introduction to this topic starting from the underlying probabilistic model, Markov Decision Process, describing how to learn a good policy (how to pick the actions) when this model is known and when it is unknown. I will stress the impact of the (required) parametrization of the solution, as well as the importance of understanding the inner engine (stochastic approximation).

I will illustrate the variety of questions by describing briefly three different questions:

- How to apply Reinforcement Learning to detect faster an issue during an ultrasound exam ?
- How to solve faster an MDP using better approximation ?
- How to make RL more robust while controlling its sample complexity ?

Orateur: LE PENNEC, Erwan (CMAP, École polytechnique, Institut Polytechnique de Paris)

ID de Contribution: 2

Type: **Non spécifié**

Learning with Missing Values: Theoretical Insights and Application to Health Databases

mercredi 3 avril 2024 11:20 (50 minutes)

Missing values are ubiquitous in many fields such as health, business or social sciences. To date, much of the literature on missing values has focused on imputation as well as inference with incomplete data. In contrast, supervised learning in the presence of missing values has received little attention. In this talk I will explain the challenges posed by missing values in regression and classification tasks. In practice, a common solution consists in imputing the missing values prior to learning. I will show how different baseline methods for handling missing values compare on several large health databases with naturally occurring missing values. We will then examine the theoretical foundations of Impute-then-Regress approaches. Finally, I will present a neural network architecture for learning with missing values that goes beyond the two-stage Impute-then-Regress approaches.

Orateur: LE MORVAN, Marine (INRIA, Saclay)

ID de Contribution: 3

Type: **Non spécifié**

Unsupervised Alignment of Graphs and Embeddings: Fundamental Limits and Computational Methods

mercredi 3 avril 2024 12:10 (50 minutes)

Aligning two (weighted or unweighted) graphs, or matching two clouds of high-dimensional embeddings, are fundamental problems in machine learning with applications across diverse domains such as natural language processing to computational biology. In this presentation I will introduce the graph alignment problem, which can be viewed as an average-case and noisy version of the graph isomorphism problem. I will talk about the main challenges when the graphs are sparse, give some insights on the fundamental limits, and present efficient algorithms for this task. Then, switching focus on aligning clouds of embeddings, I will delve into the Procrustes-Wassertein problem. We will emphasize differences from the previous graph-to-graph case. Statistical and computational results will be presented to shed light on these emerging questions.

Orateur: GANASSALI, Luca (Laboratoire de Mathématiques d'Orsay, Université Paris-Saclay)

ID de Contribution: 4

Type: **Non spécifié**

Weak Signals: Machine-Learning Meets Extreme Value Theory

mercredi 3 avril 2024 14:30 (50 minutes)

The angular measure on the unit sphere characterizes the first-order dependence structure of the components of a random vector in extreme regions and is defined in terms of standardized margins. Its statistical recovery is an important step in learning problems involving observations far away from the center. In the common situation that the components of the vector have different distributions, the rank transformation offers a convenient and robust way of standardizing data in order to build an empirical version of the angular measure based on the most extreme observations. However, the study of the sampling distribution of the resulting empirical angular measure is challenging. It is the purpose of this talk to explain how to establish finite-sample bounds for the maximal deviations between the empirical and true angular measures, uniformly over classes of Borel sets of controlled combinatorial complexity. The bounds are valid with high probability and, up to logarithmic factors, scale as the square root of the effective sample size. The bounds are next applied to provide performance guarantees for two statistical learning procedures tailored to extreme regions of the input space and built upon the empirical angular measure: binary classification in extreme regions through empirical risk minimization and unsupervised anomaly detection through minimum-volume sets of the sphere.

Orateur: CLÉMENÇON, Stephan (LTCI, Télécom Paris, Institut Polytechnique de Paris)

ID de Contribution: 5

Type: **Non spécifié**

Contextual Stochastic Bandits with Budget Constraints and Fairness Application

mercredi 3 avril 2024 15:20 (50 minutes)

We review the setting and fundamental results of contextual stochastic bandits, where at each round some vector-valued context x_t is observed and K actions are available, each action providing a stochastic reward with expectation given by some (partially unknown) function of x_t and a . The aim is to maximize the cumulative rewards obtained, or equivalently, to minimize the regret. This requires maintaining a good balance between the estimation (a.k.a., exploration) of the function and the exploitation of the estimates built. The literature also considers additional budget constraints (leading to so-called contextual bandits with knapsacks): actions now provide rewards but also costs. The literature also illustrated that costs may model fairness constraints. We will review these two lines of work and briefly describe our own contribution in this respect, related to a more direct strategy, able to handle \sqrt{T} cost constraints over T rounds, which is exactly what is needed for fairness applications. The recent results discussed at the end of the talk will be based on the joint work by Evgenii Chzhen, Christophe Giraud, Zhen Li, and Gilles Stoltz, Small total-cost constraints in contextual bandits with knapsacks, with application to fairness, Neurips, 2023.

Orateur: STOLTZ, Gilles (CNRS, LMO, Univ. Paris-Saclay)

ID de Contribution: 6

Type: **Non spécifié**

Deep Learning in Medical Imaging: The Era of Foundation Models

mercredi 3 avril 2024 16:40 (50 minutes)

Deep learning methods have a very important role in medical imaging and it had gain a lot of attention the recent years. Currently, the community is working towards the development of large deep learning models that capture complex relations of the data and can address different tasks in a holistic way. In this talk, we will discuss about recent foundation models in medical imaging and we will focus on the opportunities and challenges of such algorithms as well as recent ways to tailored them on medical data.

Orateur: VAKALOPOULOU, Maria (Centralesupélec, Université Paris-Saclay)