

8e Journée Statistique et Informatique pour la Science des Données à Paris-Saclay

Rapport sur les contributions

ID de Contribution: 1

Type: **Non spécifié**

Manifold Learning with Noisy Data

jeudi 9 mars 2023 10:00 (50 minutes)

It is a common idea that high dimensional data (or features) may lie on low dimensional support making learning easier. In this talk, I will present a very general set-up in which it is possible to recover low dimensional non-linear structures with noisy data, the noise being totally unknown and possibly large.

Then I will present minimax rates for the estimation of the support in Hausdorff distance.

Orateur: Prof. GASSIAT, Elisabeth (LMO/Université Paris-Saclay)

ID de Contribution: 2

Type: **Non spécifié**

Hybrid AI for Knowledge Representation and Model-based Image Understanding - Towards Explainability

jeudi 9 mars 2023 11:20 (50 minutes)

This presentation will focus on hybrid AI, as a step towards explainability, more specifically in the domain of spatial reasoning and image understanding. Image understanding benefits from the modeling of knowledge about both the scene observed and the objects it contains as well as their relationships. We show in this context the contribution of hybrid artificial intelligence, combining different types of formalisms and methods, and combining knowledge with data. Knowledge representation may rely on symbolic and qualitative approaches, as well as semi-qualitative ones to account for their imprecision or vagueness.

Structural information can be modeled in several formalisms, such as graphs, ontologies, logical knowledge bases, or neural networks, on which reasoning will be based. Image understanding is then expressed as a problem of spatial reasoning. These approaches will be illustrated with examples in medical imaging, illustrating the usefulness of combining several approaches.

Orateur: Prof. BLOCH, Isabelle (LIP6/Sorbonne Université - LTCI/Télécom Paris)

ID de Contribution: 3

Type: **Non spécifié**

Leveraging Knowledge to Design Machine Learning Despite the Lack of Industrial Data

jeudi 9 mars 2023 14:30 (50 minutes)

In recent years, considerable progress has been made in the implementation of decision support procedures based on machine learning methods through the exploitation of very large databases and the use of learning algorithms.

In the industrial environment, the databases available in research and development or in production are rarely so voluminous and the question arises as to whether in this context it is reasonable to use machine learning methods.

This talk presents research work around transfer learning and hybrid models that use knowledge from related application domains or physics to implement efficient models with an economy of data.

Several achievements in industrial collaborations will be presented that successfully use these learning models to design machine learning for industrial small data regimes and to develop powerful decision support tools even in cases where the initial data volume is limited.

Orateur: Prof. MOUGEOT, Mathilde (ENSIIE & Centre Borelli/ENS Paris-Saclay)

ID de Contribution: 4

Type: **Non spécifié**

Federated Learning with Communication Constraints: Challenges in Compression Based Approaches

jeudi 9 mars 2023 12:10 (50 minutes)

In this presentation, I will present some results on optimization in the context of federated learning with compression. I will first summarise the main challenges and the type of results the community has obtained, and dive into some more recent results on tradeoffs between convergence and compression rates, and user-heterogeneity. In particular, I will describe two fundamental phenomena (and related proof techniques): (1) how user-heterogeneity affects the convergence of federated optimization methods in the presence of communication constraints, and (2) the robustness of distributed stochastic algorithms to perturbation of the iterates, and the link with model compression. I will then introduce and discuss a new compression scheme based on random codebooks and unitary invariant distributions.

Orateur: Prof. DIEULEVEUT, Aymeric (CMAP/Ecole polytechnique)

ID de Contribution: 5

Type: **Non spécifié**

Covariance & Subspace Inference: Handling Robustness, Variability and Incompleteness

jeudi 9 mars 2023 15:20 (50 minutes)

In this talk, we focus on covariance matrix inference and principal component analysis in the context of non-regular data under heterogeneous environments. First, we briefly introduce mixed effects models, which are widely used to analyze repeated measures data arising in several signal processing applications that need to incorporate the same global individual's behavior with possible local variations. Then, we will expose classical strategies to learn under Gaussian assumptions. It is worth mentioning that in certain situations, in which there exist outliers within the data set, the Gaussian assumption is not valid and leads to a dramatic performance loss. To overcome this drawback, we will present an expectation-maximization-based algorithm in which the heterogeneous component is considered part of the complete data. Then, the proposed algorithm is cast into a parallel scheme, w.r.t. the individuals, in order to alleviate the computational cost and a possible central processor overload. In addition, extensions to deal with missing data, which refers to the situation where part of the individual responses is unobserved, will be presented. Finally, applications related to calibration and imaging in the context of large radio-interferometers will be considered.

Orateur: Prof. EL KORSO, Mohammed Nabil (L2S/CentraleSupélec)

ID de Contribution: 6

Type: **Non spécifié**

Transfer Learning, Covariant Learning and Parallel Transport

jeudi 9 mars 2023 16:40 (50 minutes)

Transfer learning has become increasingly important in recent years, particularly because learning a new model for each task can be much more costly in terms of training examples than adapting a model learned for another task. The standard approach in neural networks is to reuse the learned representation in the first layers and to adapt the decision function performed by the last layers.

In this talk, we will revisit transfer learning. A dual algorithm of the standard approach, which adapts the representation while keeping the decision function, will be presented, as well as an algorithm for the early classification of time series. This will allow us to question the notion of bias in transfer learning as well as the cost of information and to ask ourselves which a priori assumptions are necessary to obtain guarantees on transfer learning.

We will note that reasoning by analogy and online learning are instances of transfer learning, and we will see how the notions of parallel transport and covariant physics can provide useful conceptual tools to address transfer learning.

Orateur: Prof. CORNUEJOLS, Antoine (MIA/AgroParisTech)