

InfoMaths PhD day: titles and abstracts of scientific presentations

The 17th of June 2025

Romane Giard (Institut Camille Jordan)

Titre : Optimisation sur l'espace des mesures et estimation de modèles de mélange

Résumé : On s'intéresse à la modélisation statistique d'un jeu de données. Les modèles de mélange gaussiens sont couramment utilisés. Ils permettent notamment de faire du clustering, c'est-à-dire de séparer les données en plusieurs groupes, ce qui a de nombreuses applications. L'estimation d'un modèle de mélange gaussien pose des questions théoriques et algorithmiques. Nous examinons une méthode basée sur la résolution d'un problème d'optimisation sur l'espace des mesures, et fournissons des garanties de reconstruction dans ce cadre.

François Wieckowiak (Laboratoire d'InfoRmatique en Image et Systèmes d'information and Luminess)

Title: A Multimodal Evaluation Pipeline for Mathematical Expression Recognition

Abstract: Since OpenAI's 2016 Request for Research on Image to LaTeX, printed mathematical expression recognition has advanced significantly, with newer techniques like Vision Transformers further enhancing accuracy. However, evaluations often remain limited to a small set of datasets and primarily focus on LaTeX string similarity, neglecting semantic and visual aspects. To perform a broader evaluation, we benchmarked five models representative of the evolution of Vision Transformers from 2016 to 2024 across eight datasets. Our multimodal pipeline converts predicted LaTeX into MathML, Label Graphs, and normalized images. This range of representations allows us to compute additional metrics that better capture models' syntactic, semantic, and visual accuracy. Correlation analysis revealed that BLEU and Levenshtein alone are insufficient for complete evaluations. We propose complementary metrics, including image-based Levenshtein, MathML-based scores, and graph evaluation. Finally, MathNet achieved the highest average performance across our datasets, with a 87% LaTeX edit score, thanks to its Convolutional Vision Transformers and multi-font training dataset. However, selecting the best-performing model for each dataset yields an average edit score of 89%, indicating that combining model strengths could improve performance and generalization. Our multimodal evaluation

pipeline and metrics can offer a clearer and more robust assessment of future recognition systems.

Tristan Bullion-Gauthier (Institut Camille Jordan)

Title: Affine Sobolev inequalities

Abstract: The affine Sobolev inequality is a refinement of the Sobolev inequality. Proven by Zhang in 1999, it generalises the Petty projection inequality and implies the sharp Sobolev inequality. This inequality is also invariant under unimodular transformations. In this presentation, we will discuss some of the methods used to derive sharp Sobolev inequalities and present an elementary approach to obtaining various (non-optimal) affine inequalities in the spirit of Zhang's result.

Brandon Mosqueda (Laboratoire d'Informatique en Image et Systèmes d'information)

Title: Mitigation of Sybil-based Poisoning Attacks in Permissionless Decentralized Learning

Abstract: Decentralized learning enables collaborative machine learning with enhanced privacy by allowing participants to train models locally and share updates for aggregation instead of sharing raw data. However, such systems are vulnerable to poisoning attacks that may compromise the learning process. This threat becomes even more severe when combined with sybil attacks, where adversaries contribute numerous malicious updates with minimal effort, amplifying their impact. To overcome these challenges, particularly in the permissionless setup, we propose SyDeLP, a blockchain-enabled protocol for decentralized learning. SyDeLP integrates byzantine tolerant aggregation for poisoning mitigation with a novel Verifiable Delay Puzzle (VDP) to counter sybil attacks. Honest behavior is incentivized by dynamically reducing puzzles difficulties, decreasing the computational burden for honest nodes over time. Theoretical analysis shows the resilience of poisoning and the security of VDPs against sybil attacks. Empirical evaluations conducted on two benchmark datasets across three types of poisoning attack demonstrate that SyDeLP outperforms existing solutions in terms of poisoning resilience.