AIM workshop – speakers overview

DIAMANT sessions

Thursday (~9:10 – 10:40) // chair: Steven Kelk

1. Marten van Dijk (30min)

Title: On the Convergence of Stochastic Gradient Descent

Abstract:
We will summarize a series of results on the convergence rate of SGD starting with an upper bound on the convergence rate in an asynchronous setting for strongly convex problems. The presented framework allows us to reduce the number of communication rounds in distributed SGD (or FL) for strongly convex problems, leads to a study on the tightness of the upper bound, and motivates omega-convexity which captures problems between plain convex and strongly convex. We continue studying the convergence rate of SGD for non-convex problems. We analyze shuffling type gradient methods, introduce a variance reduction method called NC-SARAH, and if time permits give some insights on recent work on a new algorithmic framework that guarantees convergence to a global solution in the non-convex case in an "over-parameterized" setting.

2. Ronald de Wolf (30min)

Title: Mathematical Aspects of Quantum Machine Learning

Abstract:
Machine learning can be enhanced by quantum computing, both by allowing quantum data and by having quantum speedups for the optimization process that finds a good model for given data. This talk will examine some mathematical aspects of both.

3. Antonios Antoniadis (30min)

Title: Learning-Augmented Algorithms for Dynamic Power Management

Abstract:
Learning augmented algorithms is an emerging area attempting to combine the effectiveness of machine learning approaches on practically relevant inputs with the performance guarantees provided by classical worst-case analysis of online algorithms. For many problems machine learning approaches can be readily applied to generate predictions about the structure of the input. An algorithm can then be designed to incorporate such predictions in its decision process and obtain improved performance as long as these predictions are sufficiently accurate. However, in some cases the predictions may be highly inaccurate and decision-making systems that are based on such predictors need to achieve a decent performance in that case too.

In this talk we demonstrate the effectiveness of this approach by designing a learning-augmented online algorithm for the problem of minimizing power consumption in systems with multiple power-saving states. During idle periods of unknown lengths, an algorithm has to choose between power-saving states of different energy consumption and wake-up costs. The presented algorithm makes decisions based on (potentially inaccurate) predicted lengths of the idle periods and its performance is near-optimal when predictions are accurate and degrades gracefully with increasing prediction error. We show a worst-case guarantee which is almost identical to the optimal classical online algorithm for the problem. A key ingredient in our approach is a new algorithm for the online ski rental problem in the learning augmented setting with tight dependence on the prediction error.

The talk is based on results that appeared in ICML'20 and NeurIPS'21 and are joint with Christian Coester, Marek Elias, Adam Polak and Bertrand Simon.

Friday (~9:00 – 10:00) // chair: Mathias Staudigl

1. Etienne de Klerk (30min)

Title: Analyzing the worst-case behavior of popular optimization algorithms used in machine learning

Abstract:
First-order (stochastic) optimization algorithms are used extensively in machine learning (ML), for example in the training of neural networks. From the mathematical perspective, it is interesting to study the worst-case performance of these algorithms for classes of objective functions that are of interest in ML. A novel technique to perform such analysis, called semidefinite programming performance estimation, has yielded new results in this area in recent years. In this talk, we will survey this approach and some of these results, with emphasis on the (stochastic) gradient method. This is joint work with Moslem Zamani and Hadi Abbaszadehpeivasti at Tilburg University, performed as part of the NWO ENW-GROOT project *OPTIMAL: Optimization for and with Machine Learning.*

2. Ilker Birbil (30min)

Title: A Scalable Rule Generation Framework for Learning

Abstract:
We introduce a new rule-based optimization method for classification and regression. The proposed method takes advantage of linear programming and column generation, and hence, is scalable to large datasets. Moreover, the method returns a set of rules along with their optimal weights indicating the importance of each rule for learning. Through assigning cost coefficients to the rules and introducing additional constraints, we show that one can also consider interpretability and fairness of the results. We test the performance of the proposed method on a collection of datasets. Our results show that a good compromise between interpretability and fairness on the one side, and accuracy on the other side, can be obtained by the proposed rule-based learning method.

STAR sessions

Thursday (~10:55 – 12:55) // chair: Evgeny Verbitskiy

4. Peter Grünwald (30min)

Title: The Unreasonable Effectiveness of Stochastic Mathematics in Artificial Intelligence

Abstract:
We review the history of interactions between mathematics and artificial intelligence, starting from the 1956 Dartmouth meeting - a meeting of mathematiciancs - at which the phrase AI was coined to the present day, highlighting the role of probability and statistics. In particular we shall explain that the role of mathematics in machine learning and causal AI is much bigger than many of us - especially in the Netherlands - appreciate, by contrasting the quite small COLT and UAI conferences with the huge Neurips conference, the main conference in machine learning. While doing so we shall highlight mathematical research on causal inference (with a strong AI flavour) currently done in the Netherlands.

(partly based on the chapter "A Tale of Two Conferences, or: the Ongoing Dance between Math and Machine Learning" in the booklet "Mathematics: key enabling technology for scientific machine learning" edited by W. Schilders)

5. Tim van Erven (30min)

Title: Statistics and Machine Learning: Towards a Closer Integration

Abstract:
Still too often statistics and machine learning are viewed as separate fields. I will make the case that this is an outdated point of view, which does not do justice to many of the exciting new research directions being pursued in the Netherlands and abroad. I will provide many examples of the close mathematical connections between statistics and machine learning, in which ideas have fruitfully crossed from one to the other, and highlight opportunities for closer future integration both in research and in education. My examples will come both from my own expertise in online sequential prediction and explainable machine learning, and from the broader statistics and machine learning communities in the Netherlands.

6. Bert Zwart (30min)

Title: Machine Learning and applied probability

Abstract:
The Dutch applied probability community has a rich history in the analysis of Markov Decision processes which continues today.
I will give an overview, focusing on current and future challenges. In addition, I will discuss challenges related to the interaction between data, decisions, and structured queueing models. Finally, I will give a personal account of challenges at the interface of machine learning and probability, such as metastability phenomena and large deviations problems.

Friday (~10:15 – 11:15) // chair: Tim van Erven

3. Jaron Sanders (20min)

Title: Analyzing dropout training in neural networks from a stochastic perspective

Abstract:
Dropout neural networks are feed-forward neural networks in which every edge is given a random {0,1}-valued filter. These neural networks have two modes of operation: in the first each edge output is multiplied by its random filter resulting in a random output, while in the second each edge output is multiplied by the expectation of its filter leading to a deterministic output. The random mode is commonly used during training, and the deterministic mode during testing and prediction.

Many variants of dropout training exist in practice. Their regularization properties have been studied extensively, giving us understanding as to why dropout training can be beneficial. However, dropout training was not studied much from the perspective of how these random algorithms behave in detail.

I will introduce you to three results that try to give insight into the following questions:
- Is the percolation problem underlying dropout training fundamentally prohibitive?
- Can dropout neural networks still be universal approximators, in spite of us replacing random filters with expectations inside of nonlinear activations functions?
- How does the act of dropping edges actually affect the convergence rate of the training algorithm?

4. Rianne de Heide (20min)

Title: Stochastic bandits

Abstract:
This talk will be about the statistical model called the stochastic multi-armed bandit model, and in this model, an agent interacts with a sequence of probability distributions called arms. The agent sequentially selects an arm and gets a sample from the associated probability distribution, and often these samples are seen as rewards, and the agent wants to maximize the obtained rewards, or, equivalently, wants to minimize a quantity called regret. This regret-minimization setting is well-studied. Today I am also going to talk about a different setting, called best-arm identification, in which the goal is to identify the arm with the highest mean in the model as fast and accurately as possible.

There are many reasons to study bandits, because they provide a simple model for sequential decision making under uncertainty, which is a task we face in many applications: clinical trials, ad placement, recommender systems, dynamic pricing, planning in games, and the bandit model can be seen as the simplest reinforcement learning problem.

5. Arnoud den Boer

Title: Cartel formation by data-driven price algorithms

Abstract:
Can price algorithms learn to form a cartel instead of compete against each other, potentially leading to higher consumer prices and lower social welfare? The question is controversial among economists and competition policy regulators. One the one hand, concerns have been expressed that self-learning price algorithms do not only make it easier to form price cartels, but also that this can be achieved within the boundaries of current antitrust legislation – raising the question whether the existing competition law needs to be adjusted to mitigate undesired algorithmic collusion. On the other hand, a number of economists believe that algorithms learning to collude is science fiction, except by using forms of signaling or communication that are already illegal, and argue that there is no need to change antitrust laws. Motivated by this discussion, I will present recent work on learning supra-competitive prices in price and assortment games. Based on joint work with Janusz Meylahn, Thomas Loots, and Ali Aouad.

GQT sessions

Thursday (~13:30 – 15:00) // chair: Christoph Brune

7. Pepijn Roos Hoefgeest (45min)

Title: An introduction to persistent homology

Abstract:
In this talk, I will give a small introduction to persistent homology, one of the core concepts in the emerging field of Topological Data Analysis. I will give examples of how it has found use in various real world application, as well as how it relates to different fields of mathematics.

8. Bram Mesland (45min)

Title: Quantum geometry in noisy data

Abstract:
The digital age has seen an exponential increase in the availability of data, exerting an ever-larger influence on all facets of society. One challenge in data analysis is the presence of noise, interfering with the extraction of relevant information from large data sets. In mathematics, there is a strong analogy between noise and the uncertainties appearing in quantum physics, where mathematical theories have been extremely effective in deriving exact predictions. In this talk I will explore the analogy between noise and the mathematics of the quantum, in particular non-commutative geometry, and sketch possible future research avenues.

Friday (~11:15 – 12:15) // chair: Bram Mesland

6. Ieke Moerdijk (1h)

Title: Shuffles of Trees

Abstract:
Motivated by constructions is algebra and topology, we present a notion of "shuffle" of two trees. Although simple to define and useful in homotopy theory and other domains, the elementary combinatorial properties seem difficult to understand. In particular, it appears to be impossible to count the number of shuffles between two trees. The main purpose of this talk is to present this problem, in the hope that some interactions with computer science and AI might lead to an answer.

NDNS+ sessions

Thursday (~15:30 – 17:00) // chair: Tristan van Leeuwen

9. Gitta Kutyniok (45min) - online

Title: Scientific Computing meets Artificial Intelligence

Abstract:
Artificial intelligence is currently leading to one breakthrough after the other, both in public life with, for instance, autonomous driving and speech recognition, and in the sciences in areas such as medical diagnostics or molecular dynamics. A similarly strong impact can currently be witnessed for scientific computing such as for solvers of inverse problems and numerical analysis of partial differential equations.

In this lecture, we will first provide an introduction into this new vibrant research area. We will then survey recent advances at the intersection of scientific computing and artificial intelligence, and finally discuss fundamental limitations of such methodologies, in particular, in terms of computability aspects.

10. Remco Duits (45min)

Title: PDE-G-CNNs: PDE-based roto-translation equivariant convolutional neural networks and applications.

Abstract:
We consider PDE-based Group Convolutional Neural Networks (PDE-G-CNNs) that generalize Group equivariant Convolutional Neural Networks (G-CNNs).
In PDE-G-CNNs a network layer is a set of PDE-solvers where geometrically meaningful PDE-coefficients become trainable weights. The underlying PDEs are morphological and linear scale space PDEs on the homogeneous space M(d)=SE(d)/H of positions and orientations, where SE(d) is the roto-translation group and H the subgroup stabilizing a reference position and orientation, and d=2,3. The PDEs provide a geometrical understanding of the network. The network is implemented by morphological convolutions with approximations to kernels solving nonlinear HJB-PDEs (for morphological α-scale spaces), and to linear convolutions solving linear PDEs (for linear α-scale spaces). In the morphological setting, the parameter α regulates soft max-pooling over Riemannian balls, whereas in the linear setting the cases α=1/2 and α=1 correspond to the Poisson and Gaussian semigroup. We prove that our practical analytic approximation kernels are accurate. In the morphological setting, we propose analytic approximations of (sub)-Riemannian balls on M(2) which carry the correct reflectional symmetries globally and we provide asymptotic error analysis. The analytic approximations allow for efficient, accurate training of fundamental neuro-geometrical association field models in the GPU-implementations of our PDE-G-CNNs. The equivariant PDE-G-CNN network implementation consists solely of linear and morphological convolutions with parameterized analytic kernels on M(d). Common mystifying nonlinearities in CNNs are now obsolete and excluded.
We present blood vessel segmentation experiments in medical images that show clear benefits of PDE-G-CNNs compared to state-of-the-art G-CNNs: increase of performance along with a huge reduction in network parameters. We also show benefits of the equivariant CNNs over normal CNNs on mitosis/cancer detection in medical images.

Friday (~13:30 – 14:30) // chair: Tristan van Leeuwen

7. Felix Lucka (20min)

Title: Deep Learning in Computational Imaging

Abstract:
Due to its remarkable success for a variety of complex image processing problems, Deep Learning is nowadays also more commonly used in the domain of computational image reconstruction and inverse problems. In this talk, we will highlight some of the challenges and potential solutions of integrating Deep Learning into computational imaging work-flows found in scientific, clinical or industrial applications using imaging modalities such as X-ray CT, Magnetic Resonance Imaging, and Ultrasound.

8. Palina Salanevich (20min)

Title: Online non-negative matrix factorization as a tool in data processing

Abstract:
Over the last few decades, the volumes of the available data have grown exponentially. This changed our perspectives on data processing and interpretation. Many of the modern data processing techniques utilize the idea of dictionary learning, that is, data representation in terms of a reduced number of extracted features, or topics. Non-negative matrix factorization (NMF) is a powerful tool for dimensionality reduction and dictionary learning tasks. Its advantage is that topics obtained using NMF are often directly interpretable, as opposed to the representations obtained by deep neural networks. Online non-negative matrix factorization (ONMF) allows to work with data in a streaming fashion and update topics with the arrival of new data samples. This often leads to more time and memory efficient data processing methods.
In this talk, we will discuss how ONMF approach can be used for audio enhancement and, if time permits, for EEG data processing.

9. Marcello Carioni (20min)

Title: Learning adversarial regularizers for inverse problems

Abstract:
In this presentation we discuss new machine learning techniques suitable to solve ill-posed inverse problems. In particular, we deal with the task of reconstructing data from a collection of noisy measurements that are typically not enough to recover the ground-truth univocally. In this context, we propose a machine learning approach based on adversarial training that, using an unpaired set of data and measurements, is able to learn an end-to-end reconstruction of the data given the measurements. Additionally, we show how to leverage the adversarial formulation of the approach to learn, as a byproduct, a regularizer for the inverse problem that encodes prior knowledge of the data. Such regularizer is used to refine the end-to-end reconstruction by solving a variational problem that jointly penalizes the regularizer and the reconstruction error for a given observation. We showcase the potential of our approach by applying it to image reconstruction tasks in computed tomography.