Patrick Finke (UU)

Title: Memorization with ReLU nets: going beyond the worst-case

Abstract:
In practice, deep neural networks are often able to easily interpolate their training data, even when the data labels are fully randomized. To understand this phenomenon, many works have aimed to quantify the memorization capacity of a neural network architecture:

the largest number of points such that the architecture can interpolate any placement of these points with any assignment of labels.

Consequently, memorization capacity cannot take any additional structure present in real-world datasets into account and therefore amounts only to a worst-case analysis.

In this talk, I will instead take an instance-specific viewpoint and ask for the following: given a dataset (with fixed labeling), at what size is a neural network able to interpolate this specific dataset? To this end, I will introduce a randomized algorithm that provably produces an interpolating ReLU neural network. The required number of neurons and weights is connected to geometric properties of the classes and their mutual arrangement. I will illustrate the effectiveness of the algorithm with numerical experiments that underpin the theoretical findings.

The talk is based on joint work with Sjoerd Dirksen and Martin Genzel.

Maximilian Kiss (CWI)

Title: 2DeteCT - A large 2D expandable, trainable, experimental computed tomography data collection for machine learning

Abstract:
Recent research in the computational imaging community largely focuses on the use of machine learning (ML) techniques for tomographic image reconstruction. Realistic experimental data, i.e. suitable projection datasets with high-quality ground truth reconstructions and/or segmentations are scarce, but highly important for the development and training of deep learning algorithms. Our open data collection fills this gap and provides the community with a versatile 2D CT dataset for machine learning. For this, a diverse mix of samples with high natural variability in both, inter and intra-sample shape and density, were scanned with a laboratory X-ray set-up. This sample mix of dried fruits and nuts was filled in a carboard tube (height 34cm, diameter 10cm) and scanned slice by slice in a fan-beam acquisition setting. More than 5000 slices consisting of 3600 projections per full orbit were scanned in two beam settings, reconstructed with highest quality, and supplemented with a ground truth segmentation. We provide the complete image reconstruction pipeline: raw projection data (sinograms), a description of the scanning geometry, pre-processing and reconstruction scripts using open software, and the reconstructed slices as well as their ground truth segmentations. Due to this, the dataset can be used not only for ML-based reconstruction or segmentation in various settings such as limited or sparse angle, but also for noise reduction for general algorithm development.

Donato Maragno (UvA)

Title: Mixed-Integer Optimization with Constraint Learning

Abstract:
We establish a broad methodological foundation for mixed-integer optimization with learned constraints. We propose an end-to-end pipeline for data-driven decision making in which constraints and objectives are directly learned from data using machine learning, and the trained models are embedded in an optimization formulation. We exploit the mixed-integer optimization-representability of many machine learning methods, including linear models, decision trees, ensembles, and multi-layer perceptrons. The consideration of multiple methods allows us to capture various underlying relationships between decisions, contextual variables, and outcomes. We also characterize a decision trust region using the convex hull of the observations, to ensure credible recommendations and avoid extrapolation. We efficiently incorporate this representation using column generation and clustering. In combination with domain-driven constraints and objective terms, the embedded models and trust region define a mixed-integer optimization problem for prescription generation. We implement this framework as a Python package (OptiCL) for practitioners. We demonstrate the method in both chemotherapy optimization and World Food Programme planning. The case studies illustrate the benefit of the framework in generating high-quality prescriptions, the value added by the trust region, the incorporation of multiple machine learning methods, and the inclusion of multiple learned constraints.

Joint work with: Holly Wiberg, Dimitris Bertsimas, S. Ilker Birbil, Dick den Hertog, Adejuyigbe Fajemisin

Paper:
https://arxiv.org/abs/2111.04469
Code:
https://github.com/hwiberg/OptiCL

Esther Julien (TUD)

Title: Machine Learning for K-adaptability in Two-stage Robust Optimization

Abstract: Two-stage robust optimization problems constitute one of the hardest optimization problem classes. One of the solution approaches to this class of problems is K-adaptability. This approach simultaneously seeks the best partitioning of the uncertainty set of scenarios into K subsets, and optimizes decisions corresponding to each of these subsets. In general case, it is solved using the K-adaptability branch-and-bound algorithm, which requires exploration of exponentially-growing solution trees. To accelerate finding high-quality solutions in such trees, we propose a machine learning-based node selection strategy. In particular, we construct a feature engineering scheme based on general two-stage robust optimization insights that allows us to train our machine learning tool on a database of resolved B&B trees, and to apply it as-is to problems of different sizes and/or types. We experimentally show that using our learned node selection strategy outperforms a vanilla, random node selection strategy when tested on problems of the same type as the training problems, also in case the K-value or the problem size differs from the training ones.

Wout de Vos (TiU)

Title: Finding Patterns in DNA by Counting Ones

Abstract:
As large-scale datasets containing genomes from healthy and diseased people are now becoming available, there is a pressing need for sophisticated techniques to analyse these data. We derive a method that finds interpretable genotype-disease relations in large genome datasets.

We focus on Boolean rules in disjunctive normal form (DNF) that constitute a binary classification from binary input data. Our method relies on a simple statistic that counts the number of ones in columns of matrices. Inspired by the genetic algorithm, our method performs a sequence of carefully-chosen matrix-multiplications, during which patterns of interest come floating to the surface.

Lucas Vogels (UvA)

Title: Bayesian Structure Learning in undirected Graphical Models: a literature review and empirical comparison

Abstract:
Graph Neural Networks (GNN's) are a class of deep learning methods that have been adapted to leverage the structure and properties of graphs and have gained incredible success. GNN's are particularly effective when the underlying graph structure is known. However, this is not always the case. Alzheimer's diagnosis for example deals with the unknown network of the brain. Bayesian Structure Learning (BSL) is a field of research that tries to infer the properties of an unknown graph using Bayesian methods. BSL is therefore pivotal to make GNN's effective on unknown graphs. This research focuses on BSL in undirected graphical models. It compares all solution methods and their performance in a simulation study and give suggestions for future research.

Michelle Sweering (CWI)

Title: A Universal Error Measure for Input Predictions Applied to Online Graph Problems

Abstract:
We introduce a novel measure for quantifying the error in input predictions. The error is based on a minimum-cost hyperedge cover in a suitably defined hypergraph and provides a general template which we apply to online graph problems. The measure captures errors due to absent predicted requests as well as unpredicted actual requests; hence, predicted and actual inputs can be of arbitrary size. We achieve refined performance guarantees for previously studied network design problems in the online-list model, such as Steiner tree and facility location. Further, we initiate the study of learning-augmented algorithms for online routing problems, such as the online traveling salesperson problem and the online dial-a-ride problem, where (transportation) requests arrive over time (online-time model). We provide a general algorithmic framework and we give error-dependent performance bounds that improve upon known worst-case barriers, when given accurate predictions, at the cost of slightly increased worst-case bounds when given predictions of arbitrary quality.

Jelle Wemmenhove (TU/e)

Title: Educational Software for Mathematical Proof Writing

Abstract:
Based on the Coq proof assistant, we have developed a program, called Waterproof, to teach first-year students how to write mathematical proofs: students type out their proofs using controlled natural language and the computer checks the logical soundness of each proofstep; students receive feedback immediately instead of having to wait multiple days before their homework is graded.

Some automated proof finding is required: in an analysis course, students do not have to justify using the algebraic manipulations learned in high school, but the proof checker still needs to see a proof that such actions are allowed.

In the future, we wish to train two artificial agents, a student and a teacher, in order to give feedback to human students which goes beyond verification of the previous proof step.

Olov Schavemaker (UU)

Title: Approximation guarantees for random neural networks

Abstract:
With neural networks becoming more and more ubiquitous in practice, the impetus to study various aspects of them rigorously has skyrocketed. As it happens, when optimizing parameters of overparametrized deep neural networks (which are commonly used in practice) using backpropagation, the first few layers are wont to remain virtually unchanged. This is because backpropagation works its way back and the network has a nimiety of parameters, so backpropagation does not need to work its way back all the way to the start to capture sufficient complexity. Since such networks perform quite well in practice, quantifying the effect of the first few effectively random layers is of great interest from both a theoretical and a practical perspective.

My research therefore focuses on random neural networks. More specifically, on the effect of random layers and how many of them are necessary or sufficient to achieve certain approximation guarantees. In a recently submitted paper, joint with Deanna Needell, Aaron Nelson, Rayan Saab and Palina Salanevich, we provided a high-probability bound on the approximation error of a random vector functional link network -- a neural network with a single hidden layer which is random (the output weights are optimized) -- in terms of the width of the hidden layer. In my pitch, I will discuss the results obtained in this paper, and then focus on my future research plans motivated by this work.

Weihao Yan (UT)

Title: Deep architectures for Gaussian processes

Abstract:
By integrating Gaussian processes (GPs) with deep neural networks (NNs), three types of deep models are dis- cussed and compared for multivariate function approximation in this work. The first model is referred to as an *NN-equivalent GP* (Lee et al., 2017), formulated based on the fact that the

prior predictive distribution of a Bayesian NN can be equivalent to GPs under certain assumptions. An NN-induced kernel is hence defined as the GP covariance. *Deep kernel learning* (Wilson et al., 2016) is considered as a second model, which embeds a deep NN architecture into the kernel of a GP and thus improves the GP's nonlinear expressive power. The third is termed a *deep GP* model (Damianou & Lawrence, 2013), in which GPs are adopted instead of NN layers to reduce computational complexity, and variational inference is utilized to approximate the pos- terior distribution of model outputs. Numerical experiments are conducted on three benchmark problems with increased input dimensionality, and the test accuracy and Kullback–Leibler divergence are compared among the three models employing the same NN topology and GP kernel. It has been discovered that the deep kernel learning model provides superior performance. In addition, we summarize the differences among the examined models and provide an explanation why the deep kernel learn- ing model produces preferred results.

Giorgio Tosti Balducci (TUD)

Title: Predicting composite structures failure with the help of quantum neural networks

Abstract:
Composite materials failure entails many physical phenomena, which make its prediction challenging from a computational standpoint. Especially in a design context, where material properties, geometry and stacking sequence are varied, the computing time required by the finite element analyses becomes prohibitive. This has motivated the idea of data-driven composite failure prediction, where a data-based model learns from relatively few numerical simulations and it can later generalize to unseen input (e.g. loading) conditions.

Even though most machine learning models are trained on classical computers, quantum computers can also learn from data. In fact, operations on quantum bits are continuous and can be tuned in a hybrid quantum-classical optimization loop. Since the model is a quantum circuit, it is evaluated on quantum hardware, while the parameters are updated classically.

Thanks to the properties of quantum mechanics, quantum models proved able to learn functions that classical machine learning models cannot learn, especially in classification tasks. However, the datasets used were 'fabricated' such to be easily classified by the quantum circuit.

As one of the first practical applications, this work explored the power of quantum machine learning models in the practical scenario of learning composite structures failure and compared it to that of classical data-driven models.

Toby van Gastelen (CWI)

Title: Structure-Preserving Machine learning: Energy-Conserving Neural Network for Turbulence Closure Modelling

Abstract:
In turbulence modelling, and more particularly in the LES framework, we are concerned with finding a suitable closure model to represent the effect of the unresolved subgrid-scales on the larger/resolved scales. In recent years, the scientific computing community has started to gravitate

towards machine learning techniques to attempt to solve this issue. However, stability and abidance by physical structure laws of the resulting closure models is still an open problem.

To resolve this, we take the discretize first and filter next approach starting out from a high-resolution reference simulation. We apply a spatial averaging filter to reduce the degrees of freedom and derive a new kinetic energy conservation condition that takes into account both the resolved and unresolved scales. We then suggest a data-driven compression to represent the subgrid-scale content on the coarse grid in order to comply with our new conservation condition. Finally, a skew-symmetric convolutional neural network architecture is introduced that can be enhanced with dissipative terms to account for viscous flows. Combined with a structure-preserving discretization this framework is used to evolve both the filtered solution and the compressed subgrid-scale representation in time in a structure-preserving fashion, yielding non-linear stability, while still allowing for backscatter.

We apply the methodology to both the viscous Burgers' equation and Korteweg-De Vries equation in 1-D and show increased accuracy and stability as compared to a standard convolutional neural network.