DY 58: Focus Session: Physics of AI – Part II (joint session SOE/DY)

Time: Friday 9:30–12:45 Location: GÖR/0226

Invited Talk DY 58.1 Fri 9:30 GÖR/0226

What can we learn from neural quantum states? — Brandon Barton[10], Juan Carrasquilla[10], Anna Dawid[9], Antoine Georges[3,6,7,8], Megan Schuyler Moss[1,2], Alev Orfi[3,4], Christopher Roth[3], Dries Sels[3,4], Anirvan Sengupta[3,5], and •Agnes Valenti[3] — [1]Perimeter Institute for Theoretical Physics, Waterloo — [2]University of Waterloo, Waterloo — [3]Flatiron Institute, New York — [4]New York University, New York — [5]Rutgers University, New Jersey — [6]Collège de France, Paris — [7]École Polytechnique, Paris — [8]Université de Genève, Genève — [9]Universiteit Leiden, The Netherlands — [10]ETH Zürich, Switzerland

Neural quantum states (NQS) provide flexible parameterizations of quantum many-body wave-functions that serve as powerful tools for the ground-state search. At the same time, NQS offer something that standard machine-learning tasks and datasets fundamentally lack: a known underlying Hamiltonian and quantum-physics tools that allow direct examination of the encoded wavefunction. This additional structure makes NQS an interesting platform for probing the behavior of classical neural networks themselves. I will first show how pruning and scaling-law phenomena change when the learning task is the quantum wavefunction itself, and link effects depend on the underlying Hamiltonian. I will then discuss generalization and double descent through the lens of quantum observables, by analyzing how NQS fail at the interpolation threshold. Finally, I will discuss how these results relate back to practical consequences for training and architecture search in the context of the ground state search for quantum many-body systems.

DY 58.2 Fri 10:00 GÖR/0226

The NN/QFT correspondence — •Ro Jefferson — Utrecht University

Exciting progress has recently been made in the study of neural networks by applying ideas and techniques from theoretical physics. In this talk, I will discuss a precise relation between quantum field theory and deep neural networks, the NN/QFT correspondence. In particular, I will go beyond the level of analogy by explicitly constructing the QFT corresponding to a class of networks encompassing both vanilla feedforward and recurrent architectures. The resulting theory closely resembles the well-studied O(N) vector model, in which the variance of the weight initializations plays the role of the 't Hooft coupling. In this framework, the Gaussian process approximation used in machine learning corresponds to a free field theory, and finite-width effects can be computed perturbatively in the ratio of depth to width, T/N. These provide corrections to the correlation length that controls the depth to which information can propagate through the network, and thereby sets the scale at which such networks are trainable by gradient descent. If time permits, I will discuss more recent work incorporating layerwise permutation symmetry. This analysis provides a non-perturbative description of networks at initialization, and opens several interesting avenues to the study of criticality in these models.

DY 58.3 Fri 10:15 GÖR/0226

Online Learning Dynamics and Neural Scaling Laws for a Perceptron Classification Problem — •Yoon Thelge, Marcel Kuhn, and Bernd Rosenow — Institute for Theoretical Physics, University of Leipzig, 04103 Leipzig, Germany

Understanding neural scaling laws and emergence of power law generalisations remains a central challenge in learning dynamics. A natural setting for analysing this behaviour is the online-learning dynamics of a perceptron trained in a teacher*student scenario, where in the thermodynamic limit, the generalisation error exhibits characteristic power-law decay. In realistic classification problems, the teacher is a discrete classifier, while standard gradient-based training requires the student to have continuous outputs. Thus, in practically relevant settings the student is necessarily mismatched to the discrete teacher, a regime that is less well understood. We study this regime for a perceptron with a sign-activation teacher and an error-function student. We derive coupled differential equations for the evolution of the relevant order parameters and verify them via numerical integration and SGD simulations. For fixed learning rates, the generalisation error converges to zero as a power-law with respect to the number of training examples with an exponent of -1/3. The onset of this asymptotic regime shifts with the learning rate, and the generalisation at the onset scales with exponent -1/2, motivating the use of learning-rate schedules to enhance the effective asymptotic decay.

DY 58.4 Fri 10:30 GÖR/0226

Power-Law Correlations in Language: Criticality vs. Hierarchical Generative Structure — •Marcel Kühn[1,2], Max Staats[1,2], and Bernd Rosenow[2] — [1]ScaDS.AI Dresden/Leipzig, Germany — [2]Institute for Theoretical Physics, University of Leipzig, 04103 Leipzig, Germany

Natural language shows power-laws beyond Zipf: the mutual information between words as a function of separation — a two-point correlation — decays approximately as a power-law, a constraint for predictive language models. In autoregressive architectures like transformers, the softmax temperature of the output controls how sharply next-word probabilities concentrate, acting as a thermodynamic knob that might tune correlations. Since phase transitions are a well-known mechanism that generate such scale-free correlations, we ask whether the observed power-law mutual information requires tuning to a critical softmax temperature. Analyzing a Markov (bigram) model, we show that, in a large-system limit, power-law mutual information emerges only at a fine-tuned critical temperature, below correlations decay exponentially. Motivated by the fact that faithful language models must go beyond bigrams and that hierarchical generative processes introducing long range interactions are more representative, we analyze an autoregressive model that perfectly emulates a specific probabilistic context-free grammar. We demonstrate that simple versions of this model preserve power-law mutual information without temperature fine-tuning, and we discuss the generality of this result for variants of the model in which deviations from the grammatical rules may occur.

DY 58.5 Fri 10:45 GÖR/0226

Dynamics of neural scaling laws in random feature regression — •Jakob Kramp[1,2], Javed Lindner[1,2], and Moritz Helias[1,2] — [1]Institute for Advanced Simulation (IAS-6), Computational and Systems Neuroscience, Jülich Research Centre, Jülich, Germany — [2]Department of Physics, RWTH Aachen University, Aachen, Germany

Training large neural networks reveals signs of universality that hold across architectures. This holds also for overparameterized networks which converge to effective descriptions in terms of Gaussian process regression. Those simplified models, already show one ingredient of universality in form of neural scaling laws. An important ingredient are power-law distributed principal component spectra of the training data.

Past work has therefore studied the dynamics of deterministic gradient flow in linear regression with and without consideration of power-law distributed spectra. Previously, dynamics of gradient flow with power law data in a type of linear random feature model were able to mimic effects of feature learning. Our work differs from the former by presenting an approach that holds for Bayesian inference on Gaussian processes obtained by stochastic Langevin training as well as for deterministic gradient flow with or without regularization by weight decay. We obtain interpretability from an effective mean-field theory that requires fewer order parameters than previous works.

15 min. break

Invited Talk DY 58.6 Fri 11:15 GÖR/0226

Creativity in generative AI — •matthieu wyart — JHU & EPFL

Is AI creative? Generative AI such as chatGPT or diffusion models can create new texts or images from a finite training set of examples. I will argue that AI can achieve this magical by learning how compose observed low-level elements into a new whole. I will discuss the type of correlations the model can exploit to do so, how many data are needed for that, and how it relates to a hierarchical construction of latent variables. The analysis is based on the introduction of synthetic languages, and comparison with experiments performed on modern AI architectures trained on real text and images.

DY 58.7 Fri 11:45 GÖR/0226

Understanding Generative Models via Interactions — •Claudia Merger[1,2,3], Alexandre Rene[2,4], Kirsten Fischer[2,3], Peter Bouss[2,3], Sandra Nestler[2,3], David Dahmen[2], Carsten

Honerkamp[3], Moritz Helias[2,3], and Sebastian Goldt[1] — [1]SISSA, Trieste, Italy — [2]Jülich Research Centre, Jülich, Germany — [3]RWTH Aachen University, Aachen, Germany — [4]University of Ottawa, Ottawa, Canada

Generative models have become remarkably powerful at reproducing complex data distributions. They can infer the characteristic statistics of a system from comparatively small datasets and even generate new, realistic samples. Yet, our understanding of what these models learn remains limited: which statistics do they capture, and how accurately? To address the first question, we translate the statistics learned by generative models into a central concept of statistical physics: interactions between degrees of freedom that describe how pairs, triplets, and higher-order groups coact to produce the observed statistics of a system. Using invertible neural networks, we extract these interactions directly from trained models, providing a microscopic description of their learned data structure. To assess how accurately these interactions are learned, we use an analytic theory of diffusion models that predicts the precision with which pairwise interactions can be inferred from finite datasets, quantifying how generalization depends on sample size, data hierarchy, and regularization. Together, these results provide a framework grounded in statistical physics to interpret and predict the behavior of modern generative models.

DY 58.8 Fri 12:00 GÖR/0226

From Kernels to Features: A Multi-Scale Adaptive Theory of Feature Learning — •Javed Lindner[1,2], Noa Rubin[5], Kirsten Fischer[1,6], David Dahmen[1], Inbar Seroussi[4], Zohar Ringel[5], Michael Krämer[3], and Moritz Helias[1,2] — [1]Institute for Advanced Simulation (IAS-6), Computational and Systems Neuroscience, Jülich Research Centre, Jülich, Germany — [2]Department of Physics, RWTH Aachen University, Aachen, Germany — [3]Institute for Theoretical Particle Physics and Cosmology, RWTH Aachen University, Aachen, Germany — [4]Department of Applied Mathematics, School of Mathematical Sciences, Tel-Aviv University, Tel-Aviv, Israel — [5]he Racah Institute of Physics, The Hebrew University of Jerusalem, Jerusalem, Israel — [6]RWTH Aachen University, Aachen, Germany

Feature learning in neural networks is crucial for their expressive power and inductive biases, moti- vating various theoretical approaches. Some ap- proaches describe network behavior after train- ing through a change in kernel scale from initial- ization, resulting in a generalization power com- parable to a Gaussian process. Conversely, in other approaches training results in the adapta- tion of the kernel to the data, involving directional changes to the kernel. The relationship and respective strengths of these two views have so far remained unresolved. This work presents a theo- retical framework of multi-scale adaptive feature learning bridging these two views. Using methods from statistical mechanics, we derive analytical expressions for network output statistics which are valid across scaling regimes and in the continuum between them.

DY 58.9 Fri 12:15 GÖR/0226

Statistical physics of deep learning: Optimal learning of a multi-layer perceptron near interpolation — Jean Barbier[1], Francesco Camilli[1], Minh-Toan Nguyen[1], Mauro Pastore[1], and •Rudy Skerk[2] — [1]The Abdus Salam International Centre for Theoretical Physics, Strada Costiera 11, 34151 Trieste, Italy — [2]International School for Advanced Studies, Via Bonomea 265, 34136 Trieste, Italy

We address a long-standing question in statistical physics by analysing the supervised learning of a multi-layer perceptron, beyond narrow models and kernel methods. Crucially, (i) the width scales with input dimension, making the model more prone to feature learning than ultra-wide networks and more expressive than narrow ones; and (ii) we work in the interpolation regime where trainable parameters and data are comparable, forcing task-specific adaptation. In a matched teacher-student setting we establish the fundamental limits for learning random deep-network targets and identify the sufficient statistics that an optimally trained network acquires as data increases. A rich phenomenology appears with multiple learning transitions: with enough data optimal performance arises via model "specialisation", yet practical algorithms can be trapped in theory-predicted suboptimal solutions. Specialisation occurs inhomogeneously across layers, propagating from shallow towards deep ones, but also across neurons in each layer. The Bayesian-optimal analysis thus clarifies how depth, nonlinearity and finite (proportional) width shape feature learning, with implications beyond this idealised setting.

DY 58.10 Fri 12:30 GÖR/0226

Phase Transitions as Rank Transitions: Connecting Data Complexity and Cascades of Phase Transitions in analytically tractable Neural Network Models — •Björn Ladewig, Ibrahim Talha Ersoy, and Karoline Wiesner — Institute of Physics and Astronomy, University of Potsdam, Germany

Tuning the L2-regularization strength in neural networks can result in a cascade of (zero-temperature) phase transitions between regimes of increasing accuracy. This phenomenology was previously numerically observed and linked to a basin structure of the error landscape formed by the underlying data [1]. At the level of analytically tractable models, we (i) establish the existence of cascades of transitions for those models, (ii) give meaning to the transitions in terms of the ordered onset of "learned eigendirections" of the underlying data distribution; and (iii) link the phase transitions and corresponding accuracy regimes to saddle points of the error landscape.

[1] I. Talha Ersoy and Karoline Wiesner. Exploring l2-phase transitions on error landscapes. In ICML, Workshop on High-dimensional Learning Dynamics 2025. https://openreview.net/forum?id=AkQNtAw09u