FM 55: Quantum & Information Science: Neural Networks, Machine Learning, and Artificial Intelligence II

Time: Wednesday 14:00–15:30 Location: 1098

Invited Talk FM 55.1 Wed 14:00 1098

Quantum Mean Embedding of Probability Distributions — •Jonas M. Kübler, Krikamol Muandet, and Bernhard Schölkopf — Max Planck Institute for Intelligent Systems, Tübingen, Germany

The kernel mean embedding of probability distributions is commonly used in machine learning as an injective mapping from distributions to functions in an infinite dimensional Hilbert space. It allows us, for example, to define a distance measure between probability distributions, called maximum mean discrepancy (MMD). In this work we propose to represent probability distributions in a pure quantum state of a system that is described by an infinite dimensional Hilbert space and prove that the representation is unique if the corresponding kernel function is c_0-universal. This is a new method for encoding classical data in a quantum state and enables us to work with an explicit representation of the mean embedding, whereas classically one can only work implicitly with an infinite dimensional Hilbert space through the use of the kernel trick. We show how this explicit representation can speed up methods that rely on inner products of mean embeddings and discuss the theoretical and experimental challenges that need to be solved in order to achieve these speedups.

FM 55.2 Wed 14:30 1098

Training Deep Neural Networks by optimizing over paths in hyperparameter space — Vlad Pushkarov[1], Jonathan Efroni[1], Mykola Maksymenko[2], and •Maciej Koch-Janusz[3] — [1]Technion, Haifa, Israel — [2]SoftServe Inc., Lviv, Ukraine — [3]ETH Zurich, Switzerland

Hyperparameter optimization is both a practical issue and an interesting theoretical problem in training of deep architectures. Despite many recent advances the most commonly used methods almost universally involve training multiple and decoupled copies of the model, in effect sampling the hyperparameter space. We show that at a negligible additional computational cost, results can be improved by sampling paths instead of points in hyperparameter space. To this end we interpret hyperparameters as controlling the level of correlated noise in the training, which can be mapped to an effective temperature. The usually independent instances of the model are then coupled and allowed to exchange their hyperparameters throughout the training using the well established parallel tempering technique of statistical physics. Each simulation corresponds then to a unique path, or history, in the joint hyperparameter/model-parameter space. We provide empirical tests of our method, in particular for dropout and learning rate optimization. We observed faster training and improved resistance to overfitting and showed a systematic decrease in the absolute validation error, improving over benchmark results.

FM 55.3 Wed 14:45 1098

Metric Gaussian Variational Inference — •Jakob Knollmüller[1,2] and Torsten Ensslin[1,2] — [1]Max-Planck Institute for Astrophysics, Garching — [2]Ludwig-Maximilian University, Munich

A variational Gaussian approximation of the posterior distribution can be an excellent way to infer posterior quantities. However, to *capture all posterior correlations the parametrization of the full covariance is required, which scales quadratic with the problem size. This scaling prohibits full-covariance approximations for large-scale problems. As a solution to this limitation we propose Metric Gaussian Variational Inference (MGVI). This procedure approximates the variational covariance such that it requires no parameters on its own and still provides reliable posterior correlations and uncertainties for all model parameters. We approximate the variational covariance with the inverse Fisher metric, a local estimate of the true posterior uncertainty. This covariance is only stored implicitly and all necessary quantities can be extracted from it by independent samples drawn from the approximating Gaussian. MGVI requires the minimization of a stochastic estimate of the Kullback-Leibler divergence only with respect to the mean of the variational Gaussian, a quantity that scales linearly with the problem size. We motivate the choice of this covariance from an information geometric perspective. We validate the method against established approaches, demonstrate its scalability into the regime over a million parameters and capability to capture posterior distributions over complex models with multiple components and strongly non-Gaussian prior distributions. (J. Knollmüller et al. (2019) ArXiv:1901.11033)

FM 55.4 Wed 15:00 1098

Vector field divergence of predictive model output as indication of phase transitions — •Frank Schäfer and Niels Lörch — Department of Physics, University of Basel, Klingelbergstrasse 82, CH-4056 Basel, Switzerland

We introduce an alternative method to identify phase boundaries in physical systems. It is based on training a predictive model such as a neural network to infer a physical system's parameters from its state. The deviation of the inferred parameters from the underlying correct parameters will be most susceptible and diverge maximally in the vicinity of phase boundaries. Therefore, peaks in the vector field divergence of the model's predictions are used as indication of phase transitions. Our method is applicable for phase diagrams of arbitrary parameter dimension and without prior information about the phases. Application to both the two-dimensional Ising model and the dissipative Kuramoto-Hopf model show promising results.

FM 55.5 Wed 15:15 1098

Optimal Renormalization Group Transformation from Information Theory — Patrick M. Lenggenhager[1], Zohar Ringel[2], Sebastian D. Huber[1], and •Maciej Koch-Janusz[1] — [1]ETH Zurich, Switzerland — [2]Hebrew University of Jerusalem, Israel

The connections between information theory, statistical physics and quantum field theory have been the focus of renewed attention. In particular, the renormalization group (RG) has been explored from this perspective. Recently, a variational algorithm employing machine learning tools to identify the relevant degrees of freedom of a statistical system by maximizing an information-theoretic quantity, the real-space mutual information (RSMI), was proposed for real-space RG. Here we investigate analytically the RG coarse-graining procedure and the renormalized Hamiltonian, which the RSMI algorithm defines. By a combination of general arguments, exact calculations and toy models we show that the RSMI coarse-graining is optimal in a sense we define. In particular, a perfect RSMI coarse-graining generically does not increase the range of a short-ranged Hamiltonian, in any dimension. For the case of the 1D Ising model we perturbatively derive the dependence of the coefficients of the renormalized Hamiltonian on the real-space mutual information retained by a generic coarse-graining procedure. We also study the dependence of the optimal coarse-graining on the prior constraints on the number and type of coarse-grained variables. We construct toy models illustrating our findings.