Special Seminar

Department of Physics

Statistical Physics Approaches to Machine Learning Problems

Dr. Thiparat Chotibut

Singapore University of Technology and Design

Friday July 20, 2018 1.00 – 2.00 PM Rm. 610 Mahamakut Bldg.

Abstract

Advances in computing hardware and availability of massive datasets have empowered machine learning (ML) algorithms with unprecedented capabilities that steer modern economic growth. To name a few, applications of ML range from superior speech and image recognition, to autonomous vehicles, to the world's best Go player (AlphaGo). While ML is widely advertised as the new electricity, the standard practice in the field is often compared to sheer alchemy. Namely, despite tremendous empirical successes, the understanding of when and why certain algorithms work is still limited. This lack of theoretical understanding holds back the systematic development of scalable and transparent algorithms.

In this talk, I will illustrate how ideas and tools from theoretical physics can complement ML empirical practice in two widely-used but not well-understood algorithms. First, I will discuss multilayer feedforward neural networks, or *deep learning*, and associated open theoretical questions. I will show that the open question concerning the optimal choice of activation functions can be answered by applying simple dimensional analysis in conjunction with the maximum entropy principle from statistical physics. The result is the optimal activation that was recently obtained via an extensive reinforcement learning search and shown to best perform on several image classification tasks among all candidate activations. In addition, the standard ReLU activation arises as a limiting case of the proposed activation, providing the physical basis of the popular but heuristic ReLU. If time permits, the geometry of rugged high-dimensional loss landscapes associated with the proposed activation function and its connection to random matrix theory will also be discussed.

Second, I will explain the multiplicative weight update algorithm, a widely used protocol in ML and theoretical computer science that is typically employed to find solutions of optimization problems. I will show that, even when applied to solve innocuous problems in algorithmic game theory, the algorithm can generate non-convergent dynamics and chaos can arise. Remarkably, the emerging chaotic attractors possess an elegant mathematical regularity; the centers of mass of *every* attractor coincide at the mixed Nash equilibrium. I will discuss the implication of this surprising result that challenges the standard practice of algorithmic game theory, and how it solves a known open problem in ergodic theory concerning the structure of chaotic attractors. Our works exemplify the fruitfulness of interdisciplinary research at the interface among computer science, physics, and mathematics.