PMI

物理，机器与智能

Geometric landscapes and phase transitions in deep neural network models

Speaker：**Riccardo Zecchina** **(Bocconi U, Italy)**

Host：**Haiping Huang**

Abstract：

Among the most surprising aspects of deep learning models are their highly overparametrized and non-convex nature. Both of these aspects are a common trait of all the deep learning models and have led to unexpected results for classical learning theory and non-convex optimization. Current deep neural networks (DNN) are composed of millions (or even billions) of connection weights and the learning process seeks to minimize a non-convex loss function that measures the number of classification errors made by the DNN. The empirical evidence shows that these highly expressive neural network models can fit the training data via simple variants of algorithms originally designed for convex optimization. Moreover, even if the learning processes are run with little control over their statistical complexity (e.g. regularisation, number of parameters, …), these models achieve unparalleled levels of prediction accuracy, contrary to what would be expected from the uniform convergence framework of classical statistical inference. In this talk, we will discuss the geometrical structure of the space of solutions (zero error configurations) in overparametrized non-convex neural networks when trained to classify patterns taken from some natural distribution. Building on statistical physics techniques for the study of disordered systems, we analyze the geometric structure of the different minima and critical points of the error loss function as the number of parameters increases and we relate this to learning performance. Of particular interest is the role of rare flat minima which are both accessible to algorithms and have good generalisation properties, on the contrary to dominating minima which are almost impossible to sample. We will show that the appearance of rare flat minima defines a phase boundary at which algorithms start to find solutions efficiently.

About speaker：

Riccardo Zecchina is a Professor of Theoretical Physics at Bocconi University, with a chair in machine learning, previously Professor at Polytechnic University of Turin. He has been the head of the Statistical Physics Group at the International Centre for Theoretical Physics in Trieste (1997-2007). He was awarded Las Onsager Prize in 2016 (with M. Mezard and G. Parisi) for groundbreaking work applying spin glass ideas to ensembles of computational problems. His current research activity is primarily focused on the study of the out-of-equilibrium theory of learning algorithms, in artificial and biological neural networks.

Time: 17:00-18:30, Dec 10, 2021

Voovmeeting/Tecentmeeting ID：814 223 455