SQP Autumn School 2021

Introduction to some basic notions of modern machine learning

Takashi Mori
RIKEN Center for Emergent Matter Science (CEMS)

Deep learning has achieved unparalleled success in various applications such as image classification, speech recognition, natural language processing, and natural science. Remarkably, in modern machine learning applications, impressive generalization performance has been observed in an *overparameterized* regime, in which the number of trainable parameters of the network greatly exceeds the number of training data samples. Theoretically, it is recognized that the generalization error shows the so called double-descent curve: the generalization error first increases but at a certain threshold starts to decrease as the number of parameters increases. It is surprising because what we have learned in traditional statistical learning theory is that such an overparameterized network exhibits poor generalization due to overfitting to the training data. To understand the success of modern machine learning, a new theoretical framework would be demanded.

My lecture consists of two parts. In the first part, I first try to give an overview of theoretical studies of modern machine learning, including my own attempts [1, 2], with emphasis on physics perspective. I will explain double-descent curve and its implications, benefits of depth of neural networks [1], and Langevin dynamics approach to learning dynamics [2].

In the second part, I show you an analytical derivation of the double-descent curve in a simple linear regression problem. I will make a brief comment on what happens in more involved situations which are relevant to modern machine learning applications.

[1] Takashi Mori and Masahito Ueda, "Is deeper better? It depends on locality of relevant features", arXiv:2005.12488

[2] Takashi Mori, Liu Ziyin, Kangqiao Liu, and Masahito Ueda, "Logarithmic landscape and power-law escape rate of SGD", arXiv:2105.09557.