Optimal Over-Parametirized Learning

Peter Binev *
University of South Carolina

May 5, 2022

Abstract

This talk is about the problem of learning an unknown function f from given data about f. The learning problem is to give an approximation $\hat{f}$ to f that predicts the values of f away from the data. The recent advances in Deep Learning suggest using over-parameterization, namely assigning more degrees of freedom to the approximation than the number of available data. We investigate the abstract learning problem in this setting.

There are numerous settings for this learning problem depending on:
(i) what additional information we have about f (known as a model class assumption);
(ii) how we measure the accuracy of how well $\hat{f}$ predicts f;
(iii) what is known about the data and data sites;
(iv) whether the data observations are polluted by noise.
A mathematical description of the optimal performance possible (the smallest possible error of recovery) is known in the presence of a model class assumption. Under standard model class assumptions, we show that a near optimal $\hat{f}$ can be found by solving a certain discrete over-parameterized optimization problem with a penalty term. Here, near optimal means that the error is bounded by a fixed constant times the optimal error. This explains the advantage of over-parameterization which is commonly used in modern machine learning. The main results of this talk prove that over-parameterized learning with an appropriate loss function gives a near optimal approximation $\hat{f}$ of the function f from which the data is collected. Quantitative bounds are given for how much over-parameterization needs to be employed and how the penalization needs to be scaled in order to guarantee a near optimal recovery of f. An extension of these results to the case where the data is polluted by additive deterministic noise is also given.

This is a joint research project with Andrea Bonito, Ronald DeVore, and Guergana Petrova from Texas A&M University. The preprint is available at `https://arxiv.org/abs/2203.15994`

*This research was supported by the ONR Contract N00014-20-1-278, the NSF Grants DMS 1720297, DMS 2038080, DMS 2110811, DMS 2134077, and the NSF-Tripods Grant CCF-1934904.