Optimal Learning of Solutions of PDEs

Peter Binev, Andrea Bonito, Ron DeVore, Guergana Petrova

This paper studies 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. 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. Although the settings for the results are very general, the featured model assumption is that f is a solution of a (parametric) PDE and belongs to a compact manifold of the possible solutions or to a unit ball of a Banach space that contains this manifold.

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, it is shown in this paper 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 paper 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.