Kernels for deep learning - with and without tricks*

Amir Globerson

Abstract

Neural networks have recently re-emerged as a powerful hypothesis class, yielding impressive empirical performance in multiple domains. However, their training is a non-convex optimization problem which poses exciting theoretical and practical challenges. Here we argue that by extending the class of neural nets, one can obtain a convex learning problem, whose practical solution relies on the evaluation of a particular kernel (i.e., the kernel "trick"). We show that in some cases this kernel can be calculated in closed form. We next turn to the case where the kernel cannot be evaluated in closed form, and introduce a sampling based algorithm for learning with the same hypothesis class. Our regret based analysis shows that the sample complexity of the sampling algorithm is similar to that of an algorithm that uses the exact kernel. Empirical evaluation shows that the method is competitive with other kernels and sampling based algorithms.

*Machine Learning External Seminar, Gatsby Unit, July 20, 2016.