On The Power of Curriculum Learning in Training Deep Networks

Daphna Weinshall, Hebrew University of Jerusalem

Training neural networks is traditionally done by providing a sequence of random mini-batches sampled uniformly from the entire training data. In contrast, in curriculum learning teaching is achieved by gradually exposing the learner to examples in a meaningful order, from easy to hard. In this lecture, I will talk about the effects of curriculum learning on the training of deep networks, and specifically CNNs trained on image recognition. Curriculum in achieved by ranking the training points using their loss at the optimal hypothesis. I will describe an algorithm where curriculum is obtained by one of two methods: transfer learning from some competitive "teacher" network, and bootstrapping. In our empirical study both methods show similar benefits in terms of increased learning speed and improved final performance on test data. I will then outline a theoretical analysis of curriculum learning, showing how it effectively modifies the optimization landscape. To this end I will define the concept of an ideal curriculum, and show that under mild conditions it does not change the corresponding global minimum of the optimization function.

If time permits, I will describe some theoretical results in the context of convex optimization. Specifically, we analyzed the contribution of curriculum learning in two convex problems - linear regression, and binary classification by hinge loss minimization. I will show that in both cases, the expected convergence rate decreases monotonically with the ideal difficulty score, in accordance with earlier empirical results. Moreover, when the ideal difficulty score is fixed, the convergence rate is monotonically increasing with respect to the loss of the current hypothesis at each point. I will discuss how these results bring to term two apparently contradicting heuristics: curriculum learning on the one hand, and hard data mining (or active learning) on the other.