Deep Learning: Optimization, Generalization and Architectures

Amir Globerson

Professor
The Blavatnik School of Computer Science
Tel Aviv University

Abstract:

Neural networks exhibit good generalization behavior in the over-parameterized regime, where the number of network parameters exceeds the number of observations. In an attempt to explain this phenomenon, we study the problem of learning a two-layer over-parameterized neural network, when the data is generated by a linearly separable function. In this case we show that SGD converges to a global minimum that has low generalization error, and its generalization error is independent of the network size. Therefore, our result clearly shows that the use of SGD for optimization both finds a global minimum, and avoids overfitting despite the high capacity of the model. This is the first theoretical demonstration that SGD can avoid overfitting, when learning over-specified neural network classifiers. I will also discuss our recent results on using deep learning for predicting complex labels (structured-prediction), showing that such networks should obey certain structural constraints. The resulting networks show excellent performance on challenging image labeling problems.