samsi
NSF·Duke·NCSU·UNC

Deep Learning Program Opening Workshop
August 12-16, 2019

SPEAKER TITLES/ABSTRACTS

Poh-Ling Loh
University of Wisconsin

"Robust Information Bottleneck"

We derive bounds for a notion of adversarial risk, designed to characterize the robustness of linear and neural network classifiers to adversarial perturbations. Specifically, we introduce a new class of function transformations with the property that the risk of the transformed functions upper-bounds the adversarial risk of the original functions. This reduces the problem of deriving bounds on the adversarial risk to the problem of deriving risk bounds using standard learning-theoretic techniques. We then derive bounds on the Rademacher complexities of the transformed function classes, obtaining error rates on the same order as the generalization error of the original function classes. We also discuss extensions of our theory to multiclass classification and regression. Finally, we provide two algorithms for optimizing the adversarial risk bounds in the linear case, and discuss connections to regularization and distributional robustness.