Keynote: An Optimal Control View of Adversarial Machine Learning

Xiaojin Zhu*

Abstract

Test-time adversarial examples, training set poisoning, reward shaping, etc.: these attacks as studied in adversarial machine learning have one thing in common: the adversary literally wants to control a machine learning system. In this talk, we will develop this connection to control theory. The resulting view allows more clarity into adversarial learning, and opens up promising research directions.

*X. Zhu is with the Department of Computer Science, University of Wisconsin-Madison, WI, USA. e-mail: jerryzhu@cs.wisc.edu

 In: Joseph Collins, Prithviraj Dasgupta, Ranjeev Mittu (eds.): Proceedings of the AAAI Fall 2018 Symposium on Adversary-Aware Learning Techniques and Trends in Cybersecurity, Arlington, VA, USA, 18-19 October, 2018, published at http://ceur-ws.org