COMPUTATIONAL RESEARCH in BOSTON and BEYOND SEMINAR

Adversarially Robust Machine Learning Through the Lens of Optimization

ALEKSANDER MADRY

Massachusetts Institute of Technology

ABSTRACT:

Machine learning and, in particular, deep learning has made tremendous progress over the last several years. In fact, many believe now that these techniques are a "silver bullet", capable of making progress on any problem they are applied to.

But can we truly rely on this toolkit?

In this talk, I will briefly survey some of the key challenges in making machine learning be dependable and secure. Our discussion will then focus on one of the most pressing issues: the widespread vulnerability of state-of-the-art deep learning models to adversarial misclassification (aka adversarial examples). I will describe a framework that enables us to reason about this vulnerability in a principled manner using the lens of (robust) optimization. This framework provides us with a unifying perspective on much of prior work on this topic as well guides development of reliable methods for training models that are resistant to adversarial misclassification.

FRIDAY, APRIL 6, 2018
12:00 PM – 1:00 PM
Building 32, Room 124
(STATA)

Pizza and beverages will be provided.

http://math.mit.edu/crib/

Massachusetts Institute of Technology