Math 278C - Optimization and Data Science Seminar

Chester Holtz
UC San Diego

Adversarial Examples & Provable Robustness

Abstract:
Modern machine learning methods (i.e. neural networks) have been very successful in tasks such as image classification and speech recognition, but have been shown to be extremely brittle to small, adversarially-chosen perturbations of their inputs. This is a critical issue in many deep learning applications (e.g. object detection, robotic perception, ranking and recommendation, etc.). In this talk, I will provide an overview of the problem of adversarial robustness, formally introduce some general principles (what we know and what we don’t know about this phenomenon), and discuss heuristic solutions (methods that appear to work in practice) and recent certification techniques (how do we provably - and efficiently - guarantee robustness?).

Host: Jiawang Nie

Wednesday, February 3, 2021
3:00 PM
Zoom Meeting ID: 982 9781 6626 Password: 278CWn21