Computer Science Seminar

Trustworthy Machine Learning: On the Preservation of Individual Privacy and Fairness

Xueru Zhang
University of Michigan

Abstract: Machine learning (ML) techniques have seen significant advances over the last decade and are playing an increasingly critical role in people's lives. While their potential societal benefits are enormous, they can also inflict great harm if not developed or used with care. In this talk, I will focus on two critical ethical issues in ML systems: fairness and privacy, and present mitigating solutions in various scenarios. ¡br¿¡br¿On the fairness front, although many fairness criteria have been proposed to measure and remedy biases in ML systems, their impact is often only studied in a static, one-shot setting. In the first part of my talk, I will present my work on evaluating the long-term impact of (fair) ML decisions on population groups that are repeatedly subject to such decisions. I will illustrate how imposing common fairness criteria intended to protect disadvantaged groups may lead to undesirable pernicious long-term consequences by exacerbating inequality. I will then discuss a number of potential mitigations.¡br¿¡br¿ On the privacy front, when ML models are trained over individuals personal data, it is critical to preserve their individual privacy while maintaining a sufficient level of model accuracy. In the second part of the talk, I will illustrate two key ideas that can be used to balance an algorithms privacy-accuracy tradeoff: (1) reuse intermediate results to reduce information leakage; and (2) improve algorithmic robustness to accommodate more randomness. I will present a randomized, privacy-preserving algorithm that leverages these ideas in the context of distributed learning. It is shown that our algorithms privacy-accuracy tradeoff can be improved significantly over existing algorithms.

Monday, February 1, 2021, 10:00 am
https://emory.zoom.us/j/92280212733

Computer Science
Emory University