Syllabus: CS 562. Advanced Topics in Security, Privacy and Machine Learning

Instructor: Bo Li (lbo@illinois.com)
TA: Zijian Huang (zijianh4@illinois.edu)

1 Objectives

After this course, students will be able to understand security and privacy vulnerabilities of machine learning models, as well as how to make the learning systems robust from various perspectives.

2 Grading

Criteria	**Percent of Grade**
Project	65%
(Initial Proposal, Due 10.02)	(5%)
(Status Report, Due 10.26)	(20%)
(Final Report & Presentation, Due 12.05)	(40%)
Paper reading and presentation	30%
(Paper reviews)	(10%)
(Presentation)	(15%)
(Peer rating)	(5%)
Class participation	5%

Note: The presentation is evaluated based on both the content of slides and quality of presentation.

3 Prerequisites

1. All enrolled students must have taken machine learning classes.
2. Projects will require training neural networks with standard automatic differentiation packages (TensorFlow, Pytorch).
3. Tentative Goal: Everyone group in the class should have one top-tier conference paper for your project!

4 Candidate topics for final projects:

1. Attacks against general machine learning models such as 3D reconstruction, BERT, and RL systems.
2. Detection against evasion attacks such as Deepfake.
3. Robustness against poisoning attacks.
4. GWAS for AI to improve robustness

5. Certified robustness for DNNs against L_p bounded attacks
6. Certified robustness for classifiers against different types of perturbation
7. Theoretical understanding of generative models from the game theoretic perspective
8. Applications of GANs (GAN Zoo)
9. Differential private graphs, and robust graph neural networks: a) privacy attacks on GCNs, b) DP GCNs
10. Privacy analysis for generative models
11. Robust reinforcement learning
12. Improve model robustness with unlabeled data via semi-supervised learning and distributional robustness
13. Robustness testing for different deep neural networks architectures
14. Robust autoML
15. Safety-critical scenario generation for autonomous driving with Carla

5 Reading Materials

Checkout: `https://aisecure.github.io/TEACHING/CS562/CS562.html`