CS 329T: Trustworthy Machine Learning

Lab 1 : April 1, 2021

Outline

- Course Objective
- Logistics
- Pre-Lab 1 Discussion
- HW 1 overview
- Lab section preferences
- Time Zones
- Attendance

Course Overview

CS 329T is all about...

Understanding how to assess and improve trustworthiness of ML Models

Course Modules:

~ Pre-req/review: ML Fundamentals

~ Explainability

~ Fairness

~ Privacy

~ Robustness

Logistics

- Lectures/"Fireside chats": Tue 2:30pm-3:50pm PST
- Labs (attendance required in your assigned session):
  - Session 1/2: Thu 2:30pm-3:30pm PST
  - Session 3/4: TBD (based on your input)
- Web page: http://web.stanford.edu/class/cs329t/
- Gradescope (Entry code: 5VNVV4)
- Canvas (grades)
- Piazza (announcements, for all other communication)
- Late Day Policy: 5 late days with no more than 2 late days/HW
- Stanford Honor Code

Pre Lab 1 check-off

- Please upload your completed Pre-lab notebooks on Gradescope
- Pre Lab 1 addresses basic programming libraries for ML: Numpy, Pandas and Keras
- Queries for Pre Lab 1?

HW 1 overview

- Homework 1 is designed to make sure you are comfortable with ML fundamentals that will be needed in this course.

- Learning outcomes:
  - Background checkpoint
  - Introduction to Linear Model Applications

- Content:
  - Keras
  - Scikit-learn
  - numpy
  - Explanations
  - Model Stealing
  - Adversarial Attacks

Lab Section Preferences

- We will be having 4 lab sections spread across the week.
- Each student attends one lab section per week
- Tentative lab section hours:
  - Thurs 2:30-3:30pm (Shreya)
  - Thurs 2:30-3:30pm (Soham)
  - Wed 12-1pm (Shreya)
  - Fri 11-12pm (Soham)
- We have made four groups on Canvas, please sign-up for one of the sections (Canvas->people->Groups)

Time Zones

Please tell us where you are in the world
https://forms.gle/CNvKawGg59SuxPB27