Refining Adversarial Training Methods Using Game Theory
15-400 Project Milestone Report

Akhil Nadigatla

02 March 2021

1 Major Changes

Since my last update, there have been no major changes to our proposed schedule or to the project since the last meeting.

2 Accomplishments

Following the previous milestone, I have dedicated a substantial amount of time understanding various forms of adversarial attacks (e.g. evasion, poisoning, exploratory) and the mechanisms that permit these, as well as the defense mechanisms proposed for these in the literature. I believe understanding the problem in detail will be important when it comes to the 'problem-solving' stage later on in this endeavor.

3 Meeting Second Milestone

The first half of this second milestone period was spent on catching up on the previous milestones' goals i.e. running models trained on contemporary adversarial risk algorithms. The second half was, again, a literature review and study to better grasp the context of the problem and the solutions proposed.

4 Surprises

No major surprises were encountered. As aforementioned, there exists a breadth of research on adversarial risk and training; the advancements proposed and the detail to which they have been realized is still fascinating.

5 Looking Ahead

The next big part of the project will be diving into the game theoretic aspect of the project. While I have some economic background on the topic, it is nowhere near the detail required for a task of this technical detail. Hence, I am anticipating some reading on my part of game theory and popular algorithms used in that space.

6 Revisions to Future Milestones

No major revisions are necessary to future milestones at the moment. As we get into the meat of the project, however, I am anticipating a slowdown in our progress. Hence, adjustments may be necessary, but will be accordingly documented and handled in following milestone updates.

7 Resources Needed

No additional resources are needed on my end.