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

Akhil Nadigatla

14 February 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

A big bulk of the work I have done so far is getting acquainted with background information for the project. This has involved getting a better understanding of the statistical foundations behind adversarial risk and popular adversarial training algorithms. Given that I have taken not taken any MLD classes, there has been a lot of 'catching-up' to do, especially in the realm of optimization.

3 Meeting First Milestone

While I have been able to conduct a few tests on models trained around the MNIST and CIFAR-10 datasets, these have mainly been using resources provided by the Madry Lab at MIT (a pioneer research group in this space). Therefore, their effects and workings are well-documented. This has definitely improved my knowledge of the problem, but I have not been able to make any new discoveries around the less-than-expected adversarial robustness exhibited by these methods.

4 Surprises

As previously mentioned, I am blown away by the sheer volume of information and theory surrounding this subject. While it has proven to be an exciting learning opportunity, it also translates to progress being made slower than expected. Regardless, I am looking forward to getting into the nitty-gritty of adversarial training algorithms and game theory soon.

5 Looking Ahead

The PhD student that is my immediate mentor for this project, Arun Suggala, and I have a meeting scheduled soon, where we hope to schedule a regular meeting time and through which I hope to get my next steps in venturing deeper into adversarial risk.

6 Revisions to Future Milestones

No major revisions are necessary to future milestones. My caution when creating these milestones has proven helpful in ensuring that our progress is slow but steady.

7 Resources Needed

No additional resources are needed on my end.