Refining Adversarial Training Methods Using Game Theory
15-300 Project Milestone Report

Akhil Nadigatla

13 December 2020

1 Major Changes

There are no markedly major changes to our initial project proposal. However, we came to better recognize the computational hurdle associated with training models on larger datasets like ImageNet. While it was already placed in the 125% goals of the project, scaling up our techniques for such datasets will take last priority and be performed only if time permits.

2 Accomplishments

So far, I have been able to get a better understanding of the different adversarial training algorithm at play currently, including FGSM (Fast Gradient Sign Method) and PGD (Projected Gradient Descent). To get an even better understanding of the deeper theoretical foundations behind these algorithms, I also have studied some convex optimization concepts from the course website of 10-725, a course offered at CMU [1]. I also have experimented with some models and code around adversarial robustness [2].

3 Meeting First Milestone

I have accomplished a majority of what I intended from my first milestone. One thing I had included under the first milestone as an 'if time permits' goal was to train some models on the MNIST and CIFAR-10 datasets to observe the adversarial accuracy and compare them to estimates provided in the literature. I was not able to achieve this due to a couple of reasons: (1) reading on the background information to the project took longer than anticipated and, (2) setting up for remote access to the Machine Learning Department's powerful GPU machines took some time.

4 Surprises

As aforementioned, the only surprise I have faced so far is the depth of research and theory behind existing adversarial training algorithms. I do not see this as a 'bad' surprise, rather as a learning opportunity to get better acquainted with the project material.

5 Revisions to Milestones

The only revision to be made is that the training of models using existing adversarial methods will be moved to the second milestone. However, as of now, this will not change our goals for the second milestone, because I hope to do some work over winter break as well (which will definitely clear this 'backlog').

6 Resources Needed

As of now, I have all the resources I need to complete my 15-400 project.

[1] `https://www.stat.cmu.edu/~ryantibs/convexopt/`

[2] `https://adversarial-ml-tutorial.org/`