DEPENDABLE AND SECURE AI-ML (AI60006)
Assignment 1
Submission Deadline: 28-2-22

Aim

- Getting hands-on introduction to adversarial robustness in deep learning.
- Demo code is uploaded in assignment drive (install required libraries from the given link below)

  https://anaconda.org/pytorch/pytorch

  https://adversarial-ml-tutorial.org/introduction/

Problem Description:

You have been given tutorials in Jupyter Notebooks for adversarial robustness in deep learning.

Open introduction.ipynb notebook. In this notebook pre-trained ResNet50 model has been demonstrated.

Your work is to generate:

1. Random noise
2. Random noise with zero mean and unit variance
3. Strat with height and width of image shown in demo and write the predicted label with height, width, colour.

Add noise to the data to get perturbation in the input. Keep increasing the magnitude of noise in the data; mention the step size you are using. Classify the image and calculate the loss in each iteration. Show after how much perturbation the model starts to miss classify the input data.

Plot a graph to show the reduction in the image's True class probability.

Perform the same operations as mentioned above on Logistic Regression and Decision Tree Classifier.

Prepare a report of your understanding. Create a zip file named <YOUR_ROLLNO>_A1.zip and upload it on Microsoft Teams.

Link for ART toolbox is

https://github.com/Trusted-AI/adversarial-robustness-toolbox