Project Assessment

Criteria

	Task1	Task2 Opt1	Task2 Opt2	Task2 Opt4
Assignment	5%	5%	5%	5%
Background	15%	15%	15%	15%
Experimental Settings	30%	30%	20%	25%
Results and Discussions	30%	30%	25%	30%
Citation and Contribution	10%	10%	5%	5%
Implementation (scripts)	10%	10%	30%	20%

Grades

Task2

Grading Criteria: https://github.com/csce585-mlsystems/project-athena/blob/grade-2020/README.md
Feedback: https://github.com/csce585-mlsystems/project-athena/blob/grade-2020/Task2/doubleE/Comments.md

1. intro to the assignment. [0/5]
2. background. [0/15]
3. experiment settings. [30/30]
4. results and discussions. [30/30]
5. citations. [0/5]
6. contribution. [5/5]
7. implement. [10/10]

Bonus for your discussion: +10.

Common Issues

- Introduction
  - Intorduce the task in your own words
- Background
  - Introduce the attacks and methods in your own words with a little bit of math

Attack 1: Fast Gradient Sign Method (FGSM; Goodfellow, Shlens, & Szegedy, 2015)

This method processes adversarial examples as follows:

$$x' = x + \epsilon \cdot sign(\nabla_x J(x, y))$$

where x' is the adversarial image, J is the cost (loss) function of the target model f, ∇_x is the gradient with respect to the input x (original image) with corresponding correct output y (original label), and ϵ is the magnitude of the perturbation (the change made to the pixels).

https://github.com/Jacob-L-Vincent/project-athena/blob/master/reports/Report(1).ipynb

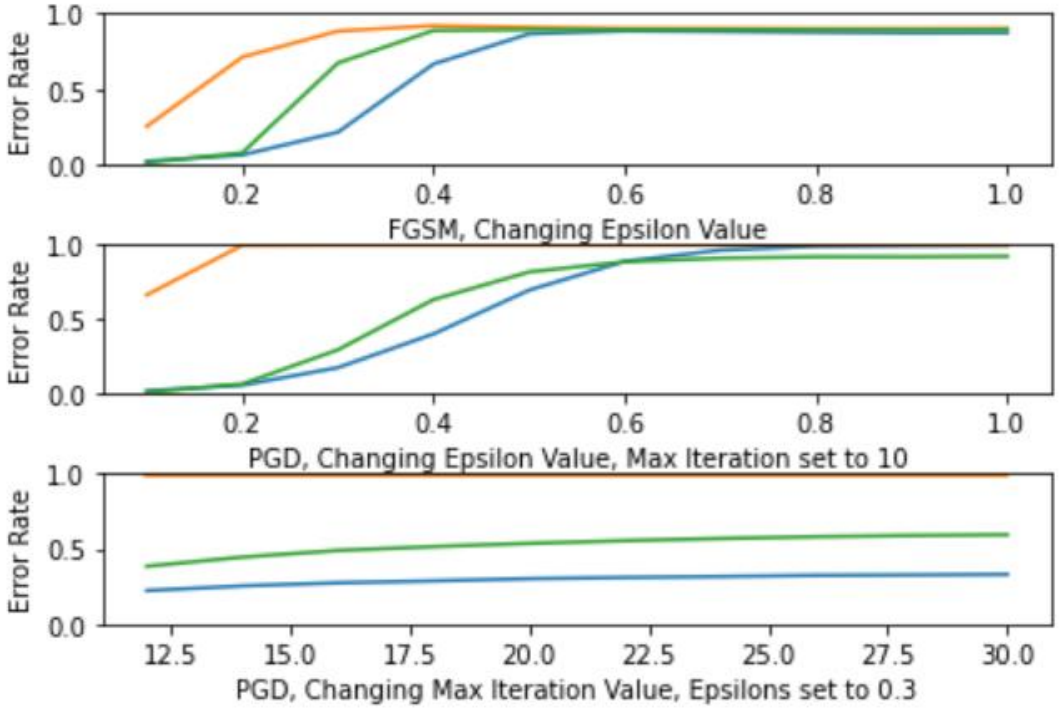
Common Issues

- Experiment
  - Not provide information regarding relevant files.
  - Ensemble size is too small: using the 3 weak defenses from demo.
  - Process
    - Subsample (if you did)
      - Ratio, location to the samples.
    - Generate AEs
      - Info of attacks: attacks, variants, path to the configuration files.
      - Info of ensemble: # weak defenses, path to the configuration files.
    - Evaluation
      - On what AEs, you evaluated what models

https://github.com/cjshearer/project-athena/blob/master/Task2/Report.ipynb

Common Issues

- Results and Discussion
  - Do not post the logs.
  - Present the results in figures or tables. Explain the result (evaluation of what, legends, etc.)
  - Observation and discussions.



https://github.com/andrewwunderlich/project-athena/blob/master/Task%201/Task1Report.pdf
https://github.com/andrewwunderlich/project-athena/blob/master/Task%202/Task2Report.pdf
https://github.com/Dojones98/project-athena/blob/master/task2/report_task2.ipynb

Common Issues

- Contribution
  - Pooyan: brainstorm; discussion; wrtiting; implement of transformations.
  - Jianhai: brainstorm; discussion; writing; implement of BB; experiment of BB.
  - Ying: brainstorm; discussion; writing; implement of framework, transformations, attacks, ZK, and WB; experiment of ZK, WB, and detector.
- Citation/Reference/Bibliograpy

References

Goodfellow, I.J., Shlens, J., & Szegedy, C. (2015). Explaining and Harnessing Adversarial Examples. CoRR, abs/1412.6572.

Madry, A., Makelov, A., Schmidt, L., Tsipras, D., & Vladu, A. (2018). Towards Deep Learning Models Resistant to Adversarial Attacks. ArXiv, abs/1706.06083.

Papernot, N., McDaniel, P., Jha, S., Fredrikson, M., Celik, Z.Y., & Swami, A. (2016). The Limitations of Deep Learning in Adversarial Settings. 2016 IEEE European Symposium on Security and Privacy (EuroS&P), 372-387.

Resubmission of deliverables

- Refine Task 1 and/or Task 2 according to the feedback.
  - Intro to task (5)
  - Background (15)
  - Citations (5)
- Due 11:59:59 PM, Dec 4th.
- Submit to folders "Task1_update" and "Task2_update" respectively.

Task 3: Presentation and Video Recording

- This is a mandatory task
- 10% (+ a bonus of 15% to the final grade)
- Due: 11:59:59 PM, Dec. 8.

Final Grade

- Task1 (30%) + Task2 (60%) + Task3 (10% + 15%)

Highlight

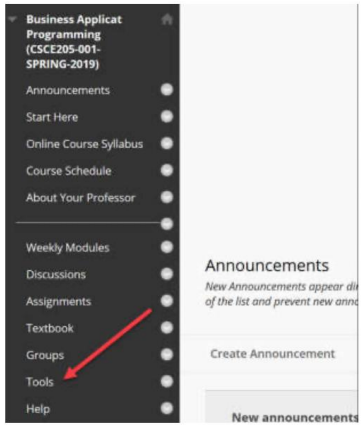
- Clutch
  - Great report of the overall framework, experiment design, and analysis.
  - Daniel Jones, Austin Staton, Ravi Patel, and Praful Chunchu
- JiR
  - Great background introduction
  - Jacob Vincent, Isaac Keohane, and Raul Ferraz
- doubleE
  - Great experiment design, analysis, and **discussion**.
  - Andrew Wunderlich, Jay Desai, and Miles Leonard-Albert
- Ares
  - Great breakdown of the task.
  - Cody Shearer, Mahmudul Hasan, Vincent Davidson, and Zhymir Thompson

Course Evaluation and Feedback

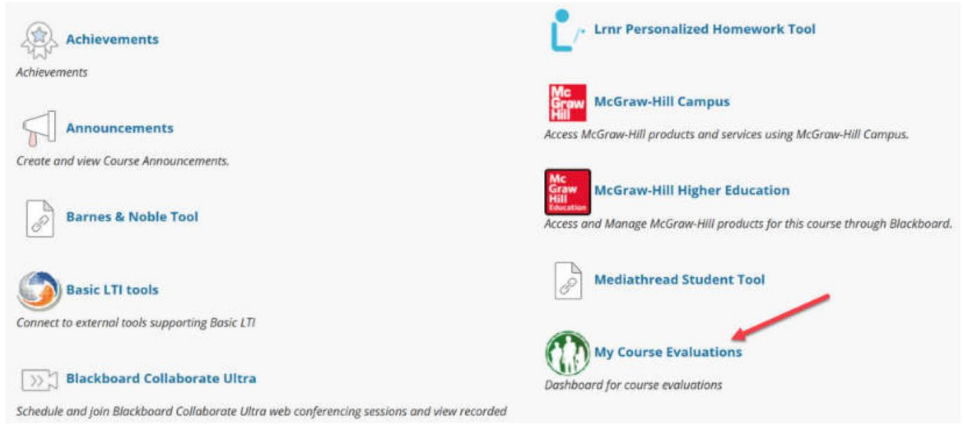
- Please complete the course evaluation in Blackboard by Dec 2nd
- I will ask also your detailed feedback by an anonymous course feedback form to be submitted by Dec 8th.

Course Evaluation Form

Once you enter your course, click **Tools**



Click on **My Course Evaluations**



From **My Survey Dashboard**, choose evaluations you wish to complete