CSCI 4963/6963 – Safe Autonomy

Course Syllabus

Course description

This course will explore the challenges with ensuring the safety of modern autonomous systems. In the first half of the course, we will discuss some of the safety vulnerabilities of machine learning models, such as neural network robustness issues, and go over methods to alleviate these issues. In the second half, we will cover concepts from dynamical system control and analysis. We will also discuss how to analyze dynamical systems with neural network controllers in the loop. Students will get exposure to adversarial learning and control techniques, as well as to a number of neural network verification tools such as Reluplex, Verisig and others.

Instructor: Rado Ivanov

Office: Lally 309
Email: ivanor@rpi.edu

Course Mechanics

Homework submission: Homework assignments will be submitted through LMS. If you have any issues with the system, let me know as soon as possible.

Attendance: While I will not formally collect attendance sheets, please try to attend all lectures, unless you have a good reason. Class participation is part of the grade, so you get free points just for attending and paying attention. Some of the assignments will be hard to complete if you don't attend the lectures.

Posting homework solutions online: You are not allowed to post your solutions on github or any other online repository. If you would like to show off your projects to potential employers, it would be much more effective to do so with extracurricular activities such as RCOS and so on.

Covid guidelines: If you have to skip a class due to covid, please let me know as soon as possible. Where possible, I will try to provide lecture recordings for those cases. **Note that lecture recordings will not be provided for any other reason and cannot be used as a substitute for in-person attendance.**

Prerequisites

Students should have a working knowledge of machine learning, probability and linear algebra. Familiarity with dynamical systems is helpful, but not required. Programming maturity in Python is assumed.

Textbooks

There is no required book for the course. All of the necessary material will be included in the lecture slides. I will suggest additional reading material before each lecture. We will follow some of the material in the following two books, which are available for free online:

- Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016. (available online: https://www.deeplearningbook.org/)
- Lee, Edward Ashford, and Sanjit A. Seshia. Introduction to embedded systems: A cyber-physical systems approach. MIT Press, 2017. (available online: https://ptolemy.berkeley.edu/books/leeseshia/)

Additionally, I will cover individual chapters from the following books. All necessary material will be provided in the lecture slides, though you are welcome to obtain your personal copies:

- Baier, Christel, and Joost-Pieter Katoen. Principles of model checking. MIT press, 2008.
- Thrun, S., Burgard, W. and Fox, D. Probabilistic robotics. Kybernetes, 2006.
- Alur, Rajeev. Principles of cyber-physical systems. MIT press, 2015.
- Scharf, Louis L., and Cédric Demeure. Statistical signal processing: detection, estimation, and time series analysis. Prentice Hall, 1991.
- Bertsekas, Dimitri. Dynamic programming and optimal control: Volume I. Vol. 1. Athena scientific, 2012.

Grading

Students will be graded on 5 homework assignments and 2 presentations. The final grade for the course will be determined as follows:

- Class participation (10%)
- Homework (60%)
- In-class presentations (30%)

Threshold	95%	90%	85%	80%	75%	70%	65%	60%	55%	50%	<50%
Grade	A	A-	B+	B	B-	C+	C	C-	D+	D	F

Late submission rule: Each homework will be worth 100 points. Late programs will be penalized 5 points per day, midnight to midnight. *Assignments which are late by more than 7 days will receive a score of 0.*

Collaboration and Academic Honesty

You are expected to work alone on all assignments. In particular:

- Discussion is allowed on homework but submitted work must be your own.
- **YOU ARE RESPONSIBLE FOR ENSURING THAT YOUR HOMEWORKS ARE NOT COPIED.**
- Copying from **anywhere** other than the class notes or your notes is NOT allowed.
- You must write and understand all solutions yourself.

In cases of academic dishonesty, the minimum penalty is a course grade of F, and other institute-mandated protocols may be invoked.