CSE 4/546 Reinforcement Learning

UB, Syllabus, Fall 2021 *

* Subject to slight modifications before the course begins

This course is intended for students interested in artificial intelligence. Reinforcement learning is an area of machine learning in which an agent learns how to behave in an environment by performing actions and assessing the results. Reinforcement learning is the method by which Google DeepMind created the AlphaGo system that beat a high-ranking Go player and how AlphaStar became the first artificially intelligent system to defeat a top professional player in StarCraft II.

Logistics

Course: Reinforcement Learning

Course number: CSE 446/546 (Senior/Graduate)

Instructor: Alina Vereshchaka

Course format: In person

Time: Tuesday and Thursday 11:10am - 12:25pm

CHECK OUR FIRST LECTURE

When? August 31, 2021 (Tuesday) 11:10am

Where? To be confirmed

Key Topics

- RL task formulation (action space, state space, environment definition). Defining RL environments.
- Tabular based solutions (dynamic programming, Monte Carlo, temporal-difference)
- Linear value function approximation
- Non-linear value function approximation (Deep Q-networks: Double DQN, Dueling DQN, PER)
- Policy gradient from basic (REINFORCE) towards advanced actor-critic algorithms (proximal policy optimization, deep deterministic policy gradient, etc.)
- Multi-agent reinforcement learning
- Imitation learning (behavioral cloning)
- Meta-learning
- Ethics & safety in AI

Pre-Requisites

CSE4/574 or CSE4/555 or CSE4/573

A few points to make sure you have the right expectations for the course so that your classroom experience will be positive.

1. All of the assignments will be completed in Python and it is assumed that you have worked with it before. Due to a busy schedule, no tutorials on Python foundations will be offered.
2. The course requires you to have prior experience working with machine learning models. It is recommended that you have taken one of our AI courses (CSE 4/573, 4/555, 4/574, or 410 Introduction to Deep Learning) or have completed a course equivalent.

3. Our second and third assignments and the final project will require us to use any of the following frameworks: Keras/PyTorch/Tensorflow. The assignment will require to build a deep learning model, so prior experience with these frameworks will be very useful.

Course Tools

To ensure we have a great experience, there is a range of tools we will be using. Try to ensure you have access to it and they are all working properly.

1. **Piazza**
   This is going to be our main method of communication. Please use Piazza for all questions related to lectures, projects, quizzes and the midterms. Unless your Piazza post needs to be private, please make the post *public*, so the answer can benefit the entire class.
2. **UBlearns**

   Weekly quizzes and assignments will be released on UBlearns. You will submit all assignments and project milestones here, where you will also find your grades.
3. **Google Calendar** (will be added later)

Assessment Overview

40% - Assignments (15% + 15% + 10%)

20% - Final Project

10% - Short Quizzes

15% - Midterm I

15% - Midterm II

GRADE DISTRIBUTION *

A [92.5,100]	B- [72.5,77.5)
A- [87.5,92.5)	C+ [67.5,72.5)
B+ [82.5,87.5)	C [62.5,67.5)
B [77.5,82.5)	C- [57.5,62.5)

* Can be adjusted before the final grade release

Assignments [40%]

We anticipate a total of **three programming assignments** throughout the course. The assignments will consist of mini programming projects, based on the topics discussed in class. A1 must be completed individually, A2 & A3 cab be completed individually or in groups of two. Assignment 1 & 2 consist of a milestone and final submissions.

These assignments are designed to give you both theoretical and practical experience with the material discussed in class. The main motivation of the assignments is for you to develop practical skills defining a problem and working with a variety of reinforcement learning algorithms.

Late Day Policy

- Students can use up to **5 free late days** throughout the course that can be applied towards the assignments (some assignments may have a hard deadline)
- A late day extends the deadline by 24 hours If there is more than 5 days after the deadline, a penalty of 25% for one day will be applied to any work submitted after that time
- For the final project there will be allocated additional late days

Final Project [20%]

The final project provides an opportunity for you to use the tools from class to build something interesting of your choice. The final project can be completed in groups of up to three. The project will be something that you work on throughout the course. Final project involves a proposal, a milestone submission, a final submission and presentation. More details will be released as the course goes.

Short Quizzes [10%]

- Each quiz contains 3-5 problems on topics covered that week
- Quizzes come in various forms, including multiple choice, multiple answer, written and coding formats
- At the end of a submission, the system will give you your final score, unless it is in the written or coding format
- 11 quizzes in total, only 10 quizzes with the highest scores will be counted
- Three attempts are allowed, unless it is in the written or coding format

Midterms [15% + 15%]

- Real-time, online over Zoom, open-book
- Questions come in various forms, including multiple choice, multiple answer, and written formats
- To ensure Academic Integrity, the midterms will be proctored over Respondus Monitor (need to have a web camera).
- Each student will have a unique set of questions with the same level of difficulty.

+ Bonuses

- **Real-Time Star** (visit ALL real-time online sessions)
- **Piazza Rockstar** (most number of endorsed and top answers on Piazza)
- **Candy Questions** (pop-up questions during real-time sessions)
- **Piazza Bonus Questions**
- **Jupyter Demo Time**
- Other activities to be included as the course goes

GPU

Over the course we will have some assignments that will require GPU access. You can use your local machine, Google Colab or you can get access to CCR. To get access to the GPU provided by CCR, you would need to create an account at CCR and send me an email.

Course Bibliography

There is no required textbook for the course. Lectures are based on materials from the following textbooks, review articles and recent papers. You will be provided with course materials to review prior to classes from the following sources along with others.

BOOKS

- Richard S. Sutton and Andrew G. Barto, "Reinforcement learning: An introduction", Second Edition, MIT Press, 2019
- Wiering, Marco, and Martijn Van Otterlo. "Reinforcement learning." Adaptation, learning, and optimization 12 (2012): 3.
- Russell, Stuart J., and Peter Norvig. "Artificial intelligence: a modern approach."Pearson Education Limited, 2016.

- Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. "Deep learning." MIT press, 2016.

SELECTED PAPERS

- Mnih, Volodymyr, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves et al. "Human-level control through deep reinforcement learning." Nature 518, no. 7540 (2015): 529.
- Van Hasselt, Hado, Arthur Guez, and David Silver. "Deep reinforcement learning with double q-learning." In Thirtieth AAAI conference on artificial intelligence. 2016.
- Wang, Ziyu, Tom Schaul, Matteo Hessel, Hado Van Hasselt, Marc Lanctot, and Nando De Freitas. "Dueling network architectures for deep reinforcement learning." arXiv preprint arXiv:1511.06581 (2015).
- Schaul, Tom, John Quan, Ioannis Antonoglou, and David Silver. "Prioritized experience replay." arXiv preprint arXiv:1511.05952 (2015).
- Bojarski, Mariusz, Davide Del Testa, Daniel Dworakowski, Bernhard Firner, Beat Flepp, Prasoon Goyal, Lawrence D. Jackel et al. "End to end learning for self-driving cars." arXiv preprint arXiv:1604.07316 (2016).
- Schulman, John, Sergey Levine, Pieter Abbeel, Michael Jordan, and Philipp Moritz. "Trust region policy optimization." In International conference on machine learning, pp. 1889-1897. 2015.
- Schulman, John, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. "Proximal policy optimization algorithms." arXiv preprint arXiv:1707.06347 (2017).
- Lillicrap, Timothy P., Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra.

"Continuous control with deep reinforcement learning." arXiv preprint arXiv:1509.02971 (2015).

Accessibility Services & Special Needs

If you have a disability and may require some type of instructional and/or examination accommodation, please inform me early in the semester so that we can coordinate the accommodations you may need. If you have not already done so, please contact the Office of Accessibility Services.

All information and documentation is confidential. The University at Buffalo and the School of Engineering and Applied Sciences are committed to ensuring equal opportunity for persons with special needs to participate in and benefit from all of its programs, services and activities.

Auditing

Auditing of the course (i.e. accessing online lectures but not turning in assignments, quizzes, projects to grade) is welcome and permitted with prior approval. Auditing will result in an 'N' grade. There is a special procedure from the department that requires students to submit a written request in order to audit the class.

For undergraduate students: Audit Form

For graduate students: please get in touch with the graduate coordinator (csegradcoord@buffalo.edu) for more details.

Academic Integrity

- Academic integrity is a fundamental university value.
- No collaboration, cheating, and plagiarism is allowed in assignments, quizzes or the midterms.

- The catalog describes plagiarism as “Copying or receiving material from any source and submitting that material as one’s own, without acknowledging and citing the particular debts to the source (quotations, paraphrases, basic ideas), or in any other manner representing the work of another as one’s own.”
- During the midterms you are supposed to be the only one in the test space. You need to inform the instructor prior to the midterm if you are planning to take a midterm in a public place, where you may encounter other people.
- Any suspicious cases will be officially reported using the Academic Dishonesty Report form and all bonus points will be subject to removal from the student’s final evaluation.
- Those found violating academic integrity more than once throughout their program will receive an immediate F in the course.
- Please refer to the Academic Integrity Policy for more details.

Questions?

Post on Piazza or send me an email to avereshc[at]buffalo.edu.

We also plan to go over course logistics and expectations in more detail during our first lecture.

FAQ

Can this course satisfy breadth/depth requirements?

Yes, the course can be used to satisfy the depth requirement for the AI focus area for graduate level (CSE 546).

I want to prepare for the course, what can I do?

There is a number of things you can do be better prepared before the course starts:

1. Request GPU resources and get familiar with CCR
2. Review Chapters 1-3 from Richard S. Sutton and Andrew G. Barto, "Reinforcement learning: An introduction", Second Edition (pdf)

What is the group size for projects?

- Assignment 1 -- individual
- Assignment 2 -- up to 2 people
- Assignment 3 -- up to 2 people

  * Teams for A2 & A3 should be different
- Final Project -- up to 3 people

I have a conflicting schedule with other courses, can I still enroll?

As part of our course we have some portion of the assessment that has to be completed real time during the lecture. Coming to the lecture 10-20 mins later may influence the performance for this assessment.