Deep Reinforcement Learning and Control
Fall 2018, CMU 10703

Instructors: Katerina Fragkiadaki, Tom Mitchell
Lectures: MW, 12:00-1:20pm, 4401 Gates and Hillman Centers (GHC)
Office Hours:

- Katerina: Tuesday 1.30-2.30pm, 8107 GHC
- Tom: Monday 1:20-1:50pm, Wednesday 1:20-1:50pm, Immediately after class, just outside the lecture room

Teaching Assistants:

- Nicholay Topin : Monday 3pm-4pm, GHC 8123
- Balarama Buddharaju : Tuesday 9:30am-10:30am, 5th floor commons
- Anshu Aviral : Tuesday 11am-12pm, 6th floor commons
- Siddharth Ancha : Wednesdays 11am-12pm, GHC 8021
- Aditya Siddhant : Thursday 5pm-6pm, LTI 5th floor commons
- Shihui Li : Thursday 10am-11am, GHC 5th floor commons
- Eric Nie : Thursday 1pm-2pm, GHC 5th floor commons
- Brynn Edmunds

Communication: Piazza is intended for all future announcements, general questions about the course, clarifications about assignments, student questions to each other, discussions about material, and so on. We strongly encourage all students to participate in discussion, ask, and answer questions through Piazza.

Class goals

- Implement and experiment with existing algorithms for learning control policies guided by reinforcement, demonstrations and intrinsic curiosity.
- Evaluate the sample complexity, generalization and generality of these algorithms.
- Be able to understand research papers in the field of robotic learning.
- Try out some ideas/extensions on your own. Particular focus on incorporating sensory input from visual sensors.

Prerequisites

The prerequisite for this course is a full semester introductory course in machine learning, such as CMU's 10-401, 10-601, 10-701 or 10-715. If you have passed a similar semester-long course at another university, we accept that. If you have not satisfied this prerequisite courses, we very strongly recommend you take the prerequisite this semester, and take 10-703 next semester.

Schedule

The following schedule is tentative, it will continuously change based on time constraints and interest of the people in the class. Reading materials and lecture notes will be added as lectures progress.

Date	*Topic (slides)*	*Lecturer*	*Readings*
08/27	Introduction	Katerina	[1, SB Ch1]
08/29	Markov decision processes (MDPs),	Katerina	[SB, Ch 3]
09/05	Solving known MDPs: Dynamic Programming, Evaluating policies	Tom	[SB, Ch 4]
09/10	Policy iteration, Value iteration, Asynchronous DP	Tom	[SB, Ch 4]
09/12	Monte Carlo Learning, Temporal difference learning, Q learning	Tom	[SB, Ch 5,6]
09/14	Recitation: OpenAI Gym recitation	Aviral	

09/17	Temporal difference learning (Tom), Planning and learning: Dyna, Monte carlo tree search (Katerina)	Katerina	[SB, Ch 8; 2;43]
09/19	Deep NN Architectures for RL	Katerina	[GBC, Ch 6, Ch 9]
09/21	Recitation on Monte Carlo Tree Search	Katerina	[2;43]
09/24	VF approximation, MC, TD with VF approximation, Control with VF approximation	Katerina	[SB, Ch 9]
09/26	Deep Q Learning : Double Q learning, replay memory	Katerina	[SB, Ch 9; 10; 44]
09/28	Recitation: Tensorflow, Keras, PyTorch Tutorial	Aviral, Balaram	
10/01	Policy Gradients I	Tom	[GBC, Ch 13]
10/03	Policy Gradients II	Tom	[GBC, Ch 13]
10/05	No Recitation		
10/08	Advanced Policy Gradients	Tom	
10/10	Evolution Methods, Natural Gradients	Katerina	
10/12	HW2 Office Hours	Shihui, Nicolay	
10/15	Natural Policy Gradients, TRPO, PPO, ACKTR	Katerina	
10/17	Pathwise Derivatives, DDPG, multigoal RL, HER	Katerina	
10/19	No Recitation		
10/22	Exploration vs. Exploitation I	Tom	[SB Ch2]
10/24	Exploration vs. Exploitation II	Katerina	
10/26	No Recitation		
10/29	Exploration and RL in Animals	Katerina, Tom	
10/31	Model-based Reinforcement Learning	Katerina	
11/02	Midway Reports Due		
11/05	Imitation Learning	Tom	
11/07	Maximum Entropy Inverse RL, Adversarial imitation learning	Katerina	
11/09	No Recitation		
11/12	Guest lecture: Research in RL	Prof. Ruslan Salakhutdinov	
11/14	Guest lecture: Experiment planning and RL	Prof. Barnabas Poczos	
11/16	Recitation: HER Practice Environment	Siddharth, Nicholay	
11/19	Poster Presentations		
11/21	Thanksgiving Break		
11/23	No Recitation		
11/26	Guest lecture: Self driving cars and RL	Prof. Jeff Schneider , Nick Rhinehart	
11/28	Maximum Entropy RL, simulation to real transfer	Katerina	
12/03	No Class		
12/05	Guest Lecture: Safe and Efficient Robot Exploration	Prof. David Held	

Resources

Readings

[SB] Sutton & Barto, *Reinforcement Learning: An Introduction*
[GBC] Goodfellow, Bengio & Courville, *Deep Learning*

[1] Smith & Gasser, *The Development of Embodied Cognition: Six Lessons from Babies*
[2] Silver, Huang et al., *Mastering the Game of Go with Deep Neural Networks and Tree Search*
[3] Houthooft et al., *VIME: Variational Information Maximizing Exploration*
[4] Stadie et al., *Incentivizing Exploration In Reinforcement Learning With Deep Predictive Models*
[5] Bagnell, *An Invitation to Imitation*
[6] Nguyen, *Imitation Learning with Recurrent Neural Networks*
[7] Bengio et al., *Scheduled Sampling for Sequence Prediction with Recurrent Neural Networks*
[8] III et al., *Searn in Practice*
[9] Bojarski et al., *End to End Learning for Self-Driving Cars*
[10] Guo et al., *Deep Learning for Real-Time Atari Game Play Using Offline Monte-Carlo Tree Search Planning*
[11] Rouhollah et al., *Learning real manipulation tasks from virtual demonstrations using LSTM*
[12] Ross et al., *Learning Monocular Reactive UAV Control in Cluttered Natural Environments*
[13] Ross et al., *A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning*
[14] Ziebart et al., *Navigate Like a Cabbie: Probabilistic Reasoning from Observed Context-Aware Behavior*
[15] Abbeel et al., *Apprenticeship Learning via Inverse Reinforcement Learning*
[16] Ho et al., *Model-Free Imitation Learning with Policy Optimization*
[17] Finn et al., *Guided Cost Learning: Deep Inverse Optimal Control via Policy Optimization*
[18] Ziebart et al., *Maximum Entropy Inverse Reinforcement Learning*
[19] Ziebart et al., *Human Behavior Modeling with Maximum Entropy Inverse Optimal Control*
[20] Finn et al., *Connection between Generative Adversarial Networks, Inverse Reinforcement Learning, and Energy-Based Models*
[21] Tassa et al., *Synthesis and Stabilization of Complex Behaviors through Online Trajectory Optimization*
[22] Watter et al., *Embed to Control: A Locally Linear Latent Dynamics Model for Control from Raw Images*
[23] Levine et al., *Learning Neural Network Policies with Guided Policy Search under Unknown Dynamics*
[24] Levine et al., *Guided Policy Search*
[25] Levine et al., *End-to-End Training of Deep Visuomotor Policies*
[26] Kumar et al., *Learning Dexterous Manipulation Policies from Experience and Imitation*
[27] Mishra et al., *Prediction and Control with Temporal Segment Models*
[28] Lillicrap et al., *Continuous control with deep reinforcement learning*
[29] Heess et al., *Learning Continuous Control Policies by Stochastic Value Gradients*
[30] Mordatch et al., *Combining model-based policy search with online model learning for control of physical humanoids*
[31] Rajeswaran et al., *EPOpt: Learning Robust Neural Network Policies Using Model Ensembles*
[32] Zoph et al., *Neural Architecture Search with Reinforcement Learning*
[33] Tzeng et al., *Adapting Deep Visuomotor Representations with Weak Pairwise Constraints*
[34] Ganin et al., *Domain-Adversarial Training of Neural Networks*
[35] Rusu et al., *Sim-to-Real Robot Learning from Pixels with Progressive Nets*
[36] Hanna et al., *Grounded Action Transformation for Robot Learning in Simulation*
[37] Christiano et al., *Transfer from Simulation to Real World through Learning Deep Inverse Dynamics Model*
[38] Xiong et al., *Supervised Descent Method and its Applications to Face Alignment*
[39] Duan et al., *One-Shot Imitation Learning*
[40] Lake et al., *Building Machines That Learn and Think Like People*
[41] Andrychowicz et al., *Learning to learn by gradient descent by gradient descent*
[42] Finn et al., *Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks*
[43] Silver et al., *Mastering the Game of Go without Human Knowledge*
[44] Mnih et al., *Playing Atari with Deep Reinforcement Learning*

General references

- Szepesvari, Algorithms for Reinforcement Learning
- Bertsekas, Dynamic Programming and Optimal Control, Vols I and II
- Puterman, Markov Decision Processes: Discrete Stochastic Dynamic Programming
- Powell, Approximate Dynamic Programming

Online courses

- Rich Sutton's class: Reinforcement Learning for Artificial Intelligence, Fall 2016
- John Schulman's and Pieter Abeel's class: Deep Reinforcement Learning, Fall 2015
- Sergey Levine's, Chelsea Finn's and John Schulman's class: Deep Reinforcement Learning, Spring 2017
- Abdeslam Boularias's class: Robot Learning Seminar
- Pieter Abeel's class: Advanced Robotics, Fall 2015
- Emo Todorov's class: Intelligent control through learning and optimization, Spring 2015
- David Silver's class: Reinforcement learning

AWS Resources

For those of you who need GPU resources, for future homeworks or the project, please read through this section carefully.

- **If you are not officially registered for this class, you are not allowed to request resources**. We will be checking before we submit requests, so please do not request access to them.
- We will be offering AWS resources. All students should join AWS educate using this link: https://aws.amazon.com/education/awseducate/ using their **@andrew.cmu.edu email address. If you do not use your andrew email address, your resources may be denied**. You should do this as soon as possible, as it can take time to set up your accounts.
- AWS NOTE: You need to back this account with your own credit/debit card and we will give out allocation codes of $50, this is important as **should you go over this $50 it will start charging to your card**, please be sure to keep an eye on your funds and not forget to terminate instances. **The university holds no responsibility in paying for additional usage**.
- We will ask you to complete an Allocation Form in order to apply for your resources. This will be made available later in the semester. Note: HW1 does not require AWS resources.

Assignments and grading

The course grade is a weighted average of assignments (60%) and a final project (40%). This year the project will be a competition on one of two or three specified topics, e.g., generalization of manipulation trajectories, or learning to navigate in mazes.
Please write all assignments in LaTeX using the NIPS style file. (sty file, tex example)

Homeworks

There will be three homework assignments for this class. The first assignment is to be completed independently. The following two assignments may be completed in teams of **two**. Only one person should submit the writeup and code on Gradescope. Additionally you should upload your code to Autolab. Please make sure **the same person who submitted the writeup and code to Gradescope is the one who submits it to Autolab.** Make sure you **mark your partner as a collaborator on Gradescope** (you do not need to do this in Autolab) and that both names are listed in the writeup. Writeups should be typeset in Latex and submitted as PDF. All code, including auxiliary scripts used for testing should be submitted with a README file to explain/document them.

Projects

Project topics will be announced a few weeks into the semester. You may work in teams of 2-4 people. Only one person should submit the project report and code on Gradescope. Additionally you should upload your code to Autolab. Please make sure **the same person who submitted the writeup and code to Gradescope is the one who submits it to Autolab.** Make sure you **mark your partner(s) as a collaborator on Gradescope** (you do not need to do this in Autolab) and that both names are listed in the writeup. Writeups should be typeset in Latex and submitted as PDF. All code, including auxiliary scripts used for testing should be submitted with a README file to explain/document them.

Grace Day/Late Homework Policy

- Homeworks: Each student has a total of 4 grace days that may be applied to the homework assignments. No more than 3 grace days may be used on any single assignment. Any assignment submitted more than 3 days past the deadline will get zero credit. Grace days will be subtracted from both students in the homework team. E.g. an assignment submitted 1 day late will result in both team members losing 1 grace day from their total allotment.
- Projects: Each team will be allotted a total of 3 grace days on the project, separate from homework grace days (unused grace days from the homework assignments **CANNOT** be applied to the project). Project late days can be on the midway and final report, but not on the poster presentation. Any project submitted more than 3 days past the deadline will get zero credit.

Course Policies

Auditing

- Official auditing of the course (i.e. taking the course for an â€œAuditâ€ grade) is not permitted this semester.
- Unofficial auditing of the course (i.e. watching the lectures online or attending them in person, but not turning in homeworks to grade) is welcome and permitted without prior approval. We give priority to students who are officially registered for the course, so informal auditors may only take a seat in the classroom is there is one available 10 minutes after the start of class. Unofficial auditors will not be given access to course materials such as homework assignments and exams.
- Please email Brynn if you need further clarification.

Extensions

In general, we do not grant extensions on assignments. There are several exceptions:

- Medical Emergencies: If you are sick and unable to complete an assignment or attend class, please go to University Health Services. For minor illnesses, we expect grace days or our late penalties to provide sufficient accommodation. For medical emergencies (e.g. prolonged hospitalization), students may request an extension afterwards and should include a note from University Health Services.
- Family/Personal Emergencies: If you have a family emergency (e.g. death in the family) or a personal emergency (e.g. mental health crisis), please contact your academic adviser or Counseling and Psychological Services (CaPS). In addition to offering support, they will reach out to the instructors for all your courses on your behalf to request an extension.
- University-Approved Absences: If you are attending an out-of-town university approved event (e.g. multi-day athletic/academic trip organized by the university), you may request an extension for the duration of the trip. You must provide confirmation of your attendance, usually from a faculty or staff organizer of the event.

For any of the above situations, you may request an extension by emailing Brynn . The email should be sent as soon as you are aware of the conflict and **at least 5 days prior to the deadline**. In the case of an emergency, no notice is needed.

Pass/Fail Policy

We allow you take the course as Pass/Fail. Instructor permission is not required. You must complete all aspects of the course (three homeworks and a project) if you take the course as Pass/Fail. What grade is the cutoff for Pass will depend on your program. Be sure to check with your program / department as to whether you can count a Pass/Fail course towards your degree requirements, notify us that you want to take the course pass/fail, and notify us of the Pass threshold your department uses (i.e., does it correspond to a grade of A, B, C, or D?)

Online and Waitlisted Students

- All lecture videos will be recorded and made available online. We are currently working with the administration to try and create an online section for this course (10-703 B) so that we can work on getting everybody off of the waitlist and officially enrolled. Please be patient, as this may take time to complete. Past experience suggests that there will be sufficient seats in the classroom for everybody who wants to take the course, so we are optimistic that all students on the waitlist are likely to be able to register within the first few weeks.
- Waitlisted students should complete all homework assignments with the rest of the class.
- The first couple lectures are likely to be quite full -- so it'd be best for waitlist students to use the livestream. We'll let you know when seats start to become available. Once that happens, you are welcome to take a **physical seat** (GHC 4401) if there is an open one **5 minutes after class has started** (e.g. 12:05pm)

Students with course conflicts

Students with timing conflicts (i.e., who have another class offered at the same time) will be permitted to take this course. However, there may be occasional days when we need for you to arrive in person during class time (e.g. for student presentations). We will let you know the dates that we require you to be available as soon as we know them.

Academic Integrity (Read this carefully!)

(Adapted from Roni Rosenfeld's 10-601 Spring 2016 Course Policies.)

Collaboration among Students

- The purpose of student collaboration is to facilitate learning, not to circumvent it. Studying the material in groups is strongly encouraged. It is also allowed to seek help from other students in understanding the material needed to solve a particular homework problem, provided no written notes (including code) are shared, or are taken at that time, and provided learning is facilitated, not circumvented. The actual solution must be done by each student alone.
- The presence or absence of any form of help or collaboration, whether given or received, must be explicitly stated and disclosed in full by all involved. Specifically, each assignment solution must include answering the following questions:
  Did you receive any help whatsoever from anyone in solving this assignment? Yes / No.
  - If you answered 'yes', give full details: _____
  - (e.g. "Jane Doe explained to me what is asked in Question 3.4")

  Did you give any help whatsoever to anyone in solving this assignment? Yes / No.
  - If you answered 'yes', give full details: _____
  - (e.g. "I pointed Joe Smith to section 2.3 since he didn't know how to proceed with Question 2")

  Did you find or come across code that implements any part of this assignment ? Yes / No. (See below policy on "found code")
  - If you answered 'yes', give full details: _____
  - (book & page, URL & location within the page, etc.).
- If you gave help after turning in your own assignment and/or after answering the questions above, you must update your answers before the assignment's deadline, if necessary by emailing the course staff.
- Collaboration without full disclosure will be handled severely, in compliance with CMU's Policy on Cheating and Plagiarism.

Previously Used Assignments

Some of the homework assignments used in this class may have been used in prior versions of this class, or in classes at other institutions, or elsewhere. Solutions to them may be, or may have been, available online, or from other people or sources. It is explicitly forbidden to use any such sources, or to consult people who have solved these problems before. It is explicitly forbidden to search for these problems or their solutions on the internet. You must solve the homework assignments completely on your own. We will be actively monitoring your compliance. Collaboration with other students who are currently taking the class is allowed, but only under the conditions stated above.

Policy Regarding "Found Code"

You are encouraged to read books and other instructional materials, both online and offline, to help you understand the concepts and algorithms taught in class. These materials may contain example code or pseudo code, which may help you better understand an algorithm or an implementation detail. However, when you implement your own solution to an assignment, you must put all materials aside, and write your code completely on your own, starting â€œfrom scratchâ€ . Specifically, you may not use any code you found or came across. If you find or come across code that implements any part of your assignment, you must disclose this fact in your collaboration statement.

Duty to Protect One's Work

Students are responsible for pro-actively protecting their work from copying and misuse by other students. If a student's work is copied by another student, the original author is also considered to be at fault and in gross violation of the course policies. It does not matter whether the author allowed the work to be copied or was merely negligent in preventing it from being copied. When overlapping work is submitted by different students, both students will be punished.
To protect future students, do not post your solutions publicly, neither during the course nor afterwards.

Penalties for Violations of Course Policies

All violations (even first one) of course policies will always be reported to the university authorities (your Department Head, Associate Dean, Dean of Student Affairs, etc.) as an official Academic Integrity Violation and will carry severe penalties.

1. The penalty for the first violation is a one-and-a-half letter grade reduction. For example, if your final letter grade for the course was to be an A-, it would become a C+
2. The penalty for the second violation is failure in the course, and can even lead to dismissal from the university.

Accommodations for Students with Disabilities

If you have a disability and have an accommodations letter from the Disability Resources office, please discuss your accommodation needs with Brynn or one of the instructors as early in the semester as possible. We will work with you to ensure that accommodations are provided as appropriate. If you suspect that you may have a disability and would benefit from accommodations but are not yet registered with the Office of Disability Resources, I encourage you to contact them at access@andrew.cmu.edu.