CS282R: Topics in Machine Learning
Inverse Problems in Reinforcement Learning

Instructor: Finale Doshi-Velez (finale@seas.harvard.edu)
TFs: Leo Benac (lbenac@g.harvard.edu) & J. Roberto Tello Ayala (jtelloayala@g.harvard.edu)
Class Time and Location: MW 9:45-11, SEC 1.402
Finale Office Hours: W 1:00-2:00pm, SEC 2.336 (starting Sept. 13)
Leo Office Hours: M 11:15-12:15 , Location TBD
Robert Office Hours: F 11:30-12:30, SEC 2.341

Overview

In the standard reinforcement learning setting, an agent learns how to optimize its rewards via interactions with the world. In this course, we will consider the flip of this question: suppose that we observe an agent acting in the world, and we know that agent is acting reasonably (that is, the agent's behavior is somehow near optimal). What does that tell us about the reward function? About the dynamics of the world?

We will first review the fundamentals through lectures, readings, and coding assignments. Students will also engage in a semester-long project applying and extending these ideas to problems related to a real healthcare scenario: decision-making in the intensive care unit (ICU). Decisions in the ICU are made by multiple different people, each of whom may have a different focus. What can we learn by observing their behavior? How can the knowledge that the clinician behavior we see is almost always a reasonable alternative be used to inform learning RL agents?

Technical Prerequisites

Prerequisites: Students are expected to be fluent in basic linear algebra (matrix manipulation), basic statistics (e.g. rules of expectations, importance sampling), and basic reinforcement learning (at a CS181 level).

Additionally, all assignments will be provided in Python, and TF support will only be provided in Python. Further, we will *not* be providing basic support for numpy, sklearn, etc. You will be expected to have the software engineering skills to work with data sets of 100,000+ rows.

Finally, you will be reading research papers, not curated notes or textbooks. This requires a level of notational and reading maturity: You must be able to manage the fact that different papers will use different notation, sometimes even different terms, for the same concept. You will also almost certainly encounter math that you are unfamiliar with. You must be willing to try to understand the main ideas and flow of an argument, even if you are not familiar with each piece; to be judicious in what you look up and what you let by (the skill of reading things where you do not understand every detail is something we will work on together; the prerequisite is being ready to engage in this type of reading).

Format, Assignments, and Assessment

The first several weeks will consist of lectures on the basics of batch reinforcement learning. Next, we will dive into more specific papers. There will be three homework assignments (24%), readings and discussion (20%), and a substantial semester-long project (56%).

We ask that you do not use LLMs to substitute for the cognitive effort of engaging with the course material as that inhibits your learning—whereas LLMs for debugging code or textual polish may enhance your productivity. We also realize this position is not practically enforceable. Regardless how you produced an output, you should be ready to explain it to the course staff without any AI or other assistance. (This includes justifying and expanding on any reading checks during classroom discussion.)

MIMIC Access

In this course, we will be giving you a cleaned up version of the MIMIC dataset, which contains ICU data from Beth Israel Hospital (overview and more documents).

It is a privilege and a responsibility to be able to work with real medical data; you must commit to taking appropriate care with this resource. Specifically, below are the instructions for getting access to this dataset. **If you do not complete your MIMIC access request by the end of Sept. 8, you will not be allowed to continue with the course.**

1. Complete certification CITI "Data or Specimens Only Research" course as an MIT affiliate (not any other institution)
   (a) Follow instructions here: https://physionet.org/about/citi-course/
   (b) The course you need to complete 'Data or Specimens Only Research' and 'Conflicts of Interest'
   (c) This course MUST be completed as an affiliate of MIT. There are similar courses called "Data and Specimens Only" at other institutions, but these will be rejected.
2. Go to https://physionet.org and create an account
3. Follow instructions at the end of https://physionet.org/content/mimiciv/2.0/ to sign the DUA and submit your CITI. Information for the form:
   Supervisor's name: Finale Doshi-Velez
   Supervisor's telephone number: (617) 384-0121
   Supervisor's email address: finale@seas.harvard.edu
   Supervisor's title: Professor (* information required for students and postdocs)
   General research area for which the data will be used:

As part of the course CS282 at Harvard, I will be exploring Inverse Reinforcement Learning methods in clinical settings. These methods can help us better understand what clinicians are optimizing for with their behavior.

4. When you done and have access to the data put your informations on this spreadsheet.

Homework

The goal of the homework assignments is to get you familiar with basic algorithms in reinforcement learning. The work that you do will provide evaluation procedures and baselines for your semester-long project. You will have two late days to use whenever you wish in the term, except for final project write-ups.

What you should submit: You should submit a write-up answering the questions posed in each assignment. Your write-up should be no more than 2 pages, though you may reference plots on additional pages (please do not make your plots tiny just to make them fit). Your code should be appended to the end of the write-up. *You will be graded on the write-up only. We will not run your code. However, not submitting your code may result in penalties to your homework grade.*

Collaboration: You must include the names of any people you worked with at the top of your write-up, and in what way you worked them (discussed ideas vs. team coding). If you code with others—which can be very productive!—you must have been an active participant. We may occasionally check in with groups to ascertain that everyone in the group was participating in a team-coding exercise. Your write-up must be your own.

Paper Discussions

Once we have completed our initial overview lectures, we will dive into current papers. For every paper, we have a reading question (see the calendar below). Before class, you will be expected to submit a *short* response to that reading question (3-4 sentences). In assessing your responses, the primary quality we will be looking for is engagement with the material, rather than regurgitation of facts. It is okay to say that you are confused or unsure, as long as you can be precise about your uncertainty.

The staff will lead the first few discussions, and then the remaining discussions for the semester will be led by the students. Each week, a team of students will be assigned as the discussion leaders.

Discussion Leaders are responsible for reading the paper in advance, answering the reading question, and leading the discussion. Leading the discussion involves

(a) Presenting a *15 minute* summary of the paper (there will be a timer). Your presentation may be slides or may be on the board.

(b) Creating discussion questions for the remainder of the class. When preparing topics, consider: How does this work compare to related work/what is the context? What are the main contributions of the work? How does the analysis support these claims? *Make sure that you are ready to help facilitate about an hour of discussion.*

You *must* have the staff review your presentation and discussion questions during office hours the week before you present. Not coming to discuss your presentation will result in a loss of participation points. The quality of your presentation will also count toward your presentation points.

Participants are responsible for reading the paper in advance, responding to the reading question, and being active participants in the discussion. Come to class prepared with either something interesting or insightful about the paper that you want to share, or a question that you would like to have clarified. Both your attendance and participation in discussion will count toward your participation points.

Semester Projects

Semester projects will be evaluated on the quality of your research process. It is entirely okay to try out a creative idea and find it doesn't pan out—as long as you can explain *why*. Relatedly, we do not want you be to encumbered by having to demonstrate a good research process and find a novel direction at the same time. It is also absolutely fine if the project ends up not being highly novel (that is, you discover later that there exists similar work).

The course staff will have a collection of suggested directions, and you may also consider other directions (must be approved by the course staff). You will work in teams of 2-3. At the end of the semester, we will discuss with teams the possibility to turn their class projects into machine learning publications.

Assessment will include three checkpoints (8% each):

- Checkpoint 1: You will submit a 2-3 paragraph summary of your intended project direction, what you will achieve by the next checkpoint, and relevant references. Be clear about (a) the question you are asking, (b) how you think you will address it, and (c) how you plan to measure success. At this stage, it is absolutely okay if your proposed approach has flaws or if your hypothesis is not correct. However, we will not be forgiving about a checkpoint that does not make a clear attempt at addressing (a)-(c) above. It is *not* sufficient to simply provide a literature review.

- Checkpoint 2: You will submit a 2 page update which includes a formal problem specification—see the elements above—and preliminary results. It is okay if your problem specification has changed from Checkpoint 1. You will be penalized for not having any initial results (which can be a preliminary exploration; it is also okay to have a set of results that convinced you to reformulate your problem specification).

- Checkpoint 3: You will submit a 2-4 page update which refines the problem specification and includes additional results.

The final report (32%) will be an expansion of this basic format. It is absolutely critical that your writing is clear, and that you explain *why* your ideas succeeded or failed. A series of indecipherable equations followed by dazzling plots alone will not result in a high score, no matter how dazzling the plots!

In terms of structure, your final report should include an introduction which lays out your motivation, challenge, hypothesis, and contribution; a precise methods section; and a results and discussion section that that provide not only your results but the new understanding that came from your project. There is no page limit: take the space to be complete and precise, do not be verbose.

Finally, at each stage, you will be given feedback. You will be able to earn points back for previous stages if you address the feedback in the next stage. (For example, if you submit Checkpoint 1 with an insufficient evaluation plan, but correct that in Checkpoint 2, those points will flow back to Checkpoint 1.)

Calendar

The following is a calendar of readings and assignments. Assignments will be due on Canvas.

Date	Topic	Readings	Reading Question, Notes
Lectures			
Wednesday, September 6	RL Basics: PI, VI, SARSA, q-Learning (tabular)	Sutton and Barto Ch. 1-6; other useful references include Model-Based Bayesian Exploration, Dearden et al.; FQI: Tree-Based Batch Mode Reinforcement Learning, Ernst et al.; PPO: Proximal Policy Optimization Algorithms, Schulman et al.	**Homework 1 out**
Monday, September 11	Imitation Learning Basics	A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning	
Wednesday, September 13	Inverse RL Basics	Algorithms for Inverse Reinforcement Learning, Maximum Entropy Inverse Reinforcement Learning; other useful references include Chapter 6.7 of this set of Lecture Notes. Also, these are readings about Max-Causal-Entropy, which is a more robust variant: A Primer on Maximum Causal Entropy Inverse Reinforcement Learning Modeling interaction via the principle of maximum causal entropy	**Homework 1 due, Homework 2 out**
Basic Methods			
Monday, September 18	Learning in Batch Settings	Truly Batch Apprenticeship Learning with Deep Successor Features	Why is it harder to do IRL in batch or offline settings?
Wednesday, September 20	Nonlinear MaxEnt	Guided cost learning: Deep inverse optimal control via policy optimization	How do max-ent type approaches to IRL compare to LP-type approaches? **Homework 2 due, Homework 3 out**
Monday, September 25	Adversarial Approaches for IRL	Learning robust rewards with adversarial inverse reinforcement learning	How do adversarial methods compare to max-ent methods?
Wednesday, September 27	Adversarial Approaches for Imitation Learning	Generative adversarial imitation learning	We have now seen state of the art approaches for both IRL and IL. When might you apply IRL? IL?
Monday, October 2 1 Checkpoint 1 Presentations		**Checkpoint 1 due**	
Wednesday, October 4	Bayesian IRL	Bayesian Inverse Reinforcement Learning	What is the main difference in how BIRL approaches non-identifiability compared to the approaches we have studied so far?
Different Kinds of Human Input			
Wednesday, October 11	Learning from Trajectory Preferences	Deep reinforcement learning from human preferences	How does this input differ from the previous IRL papers? What information does it provide that the previous papers do not?
Monday, October 16	IRL with Multiple Tasks	Repeated Inverse Reinforcement Learning	How do multiple tasks improve identifiability?
Wednesday, October 18	Checkpoint 2 Presentations		**Checkpoint 2 due**
Monday, October 23	IRL based on Human Reward Feedback	Learning Non-Myopically from Human-Generated Reward, Programming by Feedback	How does this feedback differ the standard IRL setting? What information does it provide that the previous papers do not?
Wednesday, October 25	IRL with Multiple Human Inputs	Reward learning from human preferences and demonstrations in Atari, Iterative interactive reward learning	Now, multiple types of human input are combined. How do they complement each other?
Monday, October 30	Unifying Framework for Reward Learning	Reward-rational (implicit) choice: A unifying formalism for reward learning	How does this paper make formal the ideas in the previous collection of papers on different kinds of human input?
Wednesday, November 1	How do different feedback relate theoretically?	Invariance in Policy Optimisation and Partial Identifiability in Reward Learning	How does this paper make formal the ideas in the previous collection of papers on different kinds of human input?
Monday, November 6	Checkpoint 3 Presentations		**Checkpoint 3 due**
Theoretical Grounding			
Wednesday, November 8	Overview of Identifiability in IRL	Identifiability in inverse reinforcement learning	What are the main factors that affect non-identifiability in IRL?
Monday, November 13	How well can we recover the feasible reward set?	Towards Theoretical Understanding of Inverse Reinforcement Learning	How do different properties of the environment and the human demonstrations affect our ability to characterize the set of feasible rewards?
Wednesday, November 15	What errors arise when the transition function is unknown?	Provably Efficient Learning of Transferable Rewards	How can we learn rewards that can be used in new settings? How does imperfect access to the transition function affect this transferability?

Monday, November 20	What about gamma?	On the Effective Horizon of Inverse Reinforcement Learning; other interesting paper Learning Rewards and Dynamics simultaneously	Why might it make sense to learn both the rewards and the discount factor when performing IRL?
Monday, November 27	Learning from Demonstration	Learning From Demonstration; Policy Optimization with Demonstrations	How do the ideas for learning from demonstration for optimization compare to IRL?
Wednesday, November 29	Final Project Presentations		
Wednesday, December 4	Final Project Presentations		**Final Papers Due December 6**