CS 7642: Reinforcement Learning and Decision Making **Project #1**

TD(λ)

1 Problem

1.1 Description

One aspect of research in reinforcement learning (or any scientific field) is the replication of previously published results. There are a few benefits you might reap from replicating papers. One benefit of replication is that it augments your understanding of the material. Another benefit is that it puts you in a good position both to extend existing literature and consider new contributions to your field. Replication is also often challenging. You may find that values of key parameters are missing, that described methods are ambiguous, or even that there are subtle errors. Sometimes obtaining the same pattern of results is not possible.

For this project, you will read Richard Sutton's 1988 paper "Learning to Predict by the Methods of Temporal Differences." Then you will create an implementation and replication of the results found in figures 3, 4, and 5. It might also be informative to compare these results with those in Chapter 7 of Sutton's textbook [1].

You will present your work in a written report of a maximum of 5 pages. The report should include a description of the experiments replicated, how the experiments were implemented (the environment, algorithms, etc), and the outcomes of the experiments. You should provide an analysis of these results. What exactly do the results demonstrate? Are there any significant differences between your results and the results in the original paper? How can you explain those differences? Describe any pitfalls you ran into while trying to replicate the experiment from the paper (e.g. unclear parameters, contradictory descriptions of the procedure to follow, results that differ wildly from the published results). What steps did you take to overcome those pitfalls? What assumptions did you make? And, why were these assumptions justified? Add anything else that you think is relevant to discuss.

1.2 Procedure

As noted, replicating results can be challenging. Expect some issues along the way and be prepared to resolve them.

- Read Sutton's Paper
- Write the code necessary to replicate Sutton's experiments
  - You will be replicating figures 3, 4, and 5 (Check Erratum at the end of the paper)
- Create the graphs
  - Replicate figures 3, 4, and 5
  - Graphs of anything else you may think appropriate
- We've created a private Georgia Tech GitHub repository for your code. Push your code to the personal repository found here: `https://github.gatech.edu/gt-omscs-rldm`
  - The quality of the code is not graded. You don't have to spend countless hours adding comments, etc. But, it will be examined by the TAs.
  - Make sure to include a `README.md` file for your repository
    - Include thorough and detailed instructions on how to run your source code in the `README.md`
    - If you work in a notebook, like Jupyter, include an export of your code in a `.py` file along with your notebook
    - The README.md file should be placed in the project_1 folder in your repository.

  - You will be penalized by 25 points if you:
    - Do not have any code or do not submit your full code to the GitHub repository
    - Do not include the git hash for your last commit in your paper

- Write a paper describing the experiments, how you replicated them, and any other relevant information.
  - Include the hash for your last commit to the GitHub repository in the paper's header.
  - 5 pages maximum – really, you will lose points for longer papers.
  - Make sure your graphs are legible and you cite sources properly. While it is not required, we recommend you use a conference paper format.
  - Describe the problem
    - You should assume your reader has not read Sutton 88 and provide sufficient background for them to understand your work and its significance. Don't cut corners here. We've never read your take and analysis of the random walk.
  - Your graphs
    - And, discussions regarding them
  - Describe the experiments
    - Discuss the implementation
    - Discuss the outcome
    - The generated data
  - Analyze your results
    - How do they match
    - How do they differ
    - Why is this the case and why is it important? Analyze your results in the context of the problem and the approach. Your analysis is where you demonstrate your understanding to the reader.
  - Describe any problems/pitfalls you encountered
    - How did you overcome them
    - What were your assumptions/justifications for this solution
  - Yes, it can be done within 5 pages and in normal font size
  - Save this paper in PDF format
  - Submit!
  - It is recommended that you download your PDF after submitting it to make sure it's what you think it is. There is no second chance to fix a corrupted or incorrect submission after the deadline.
  - Finally, under no circumstances should you use code or writing found online, or from another student in the current or a previous semester. Doing so is a violation of the Georgia Tech Honor Code, and is subject to strict disciplinary action.

2 Resources

2.1 Lectures

- Lesson 4: TD and Friends

2.2 Readings

- Sutton (1988) [2]
- Chapter 7 (7.1 n-step TD Prediction) and Chapter 12 (12.2 TD(λ)) of [1]

3 Submission Details

The due date is indicated on the Canvas page for this assignment. Make sure you have set your timezone in Canvas to ensure the deadline is accurate.
Due Date: **Indicated as "Due" on Canvas**
Late Due Date [**20 point penalty per day**]: **Indicated as "Until" on Canvas**
The submission consists of:

- Your written report in PDF format (Make sure to include the git hash of your last commit)
- Your source code in your personal repository on Georgia Tech's private GitHub

To complete the assignment, submit your written report to Project 1 under your Assignments on Canvas: https://gatech.instructure.com

You may submit the assignment as many times as you wish up to the due date, but, we will only consider your last submission for grading purposes. Late submissions will receive a cumulative 20 point penalty per day. That is, any project submitted after midnight AoE ("Anywhere on Earth") on the due date gets a 20 point penalty. Any project submitted after midnight AoE the following day gets a 40 point penalty and so on. To be clear, midnight means 12:00:00 a.m. AoE, so the last penalty-free second for submission is 11:59:59 p.m. AoE. No project will receive a score less than a zero no matter what the penalty. Any projects more than 4 days late and any projects not submitted will receive a 0.

Note: Late is late. It does not matter if you are 1 second, 1 minute, or 1 hour late. If Canvas marks your assignment as late, you will be penalized. So a submission that is 1 second beyond a due date receives the same penalty as one that is 23 hours, 59 minutes, and 59 seconds beyond that same due date. **Additionally, if you resubmit your project and your last submission is late, you will incur the penalty corresponding to the time of your last submission.**

Finally, if you have received an exception from the Dean of Students for a personal or medical emergency we will consider accepting your project up to 7 days after the initial due date with no penalty. Students requiring more time should consider withdrawing from the course (if possible) or taking an incomplete for this semester.

3.1 Grading and Regrading

When your projects are graded, you will receive feedback explaining your errors (and your successes!) in some level of detail. This feedback is for your benefit, both on this assignment and for future assignments. It is considered a part of your learning goals to internalize this feedback. This is one of many learning goals for this course, such as: understanding game theory, random variables, and noise.

If you are convinced that your grade is in error in light of the feedback, you may request a regrade within a week of the grade and feedback being returned to you. A regrade request is only valid if it includes an explanation of where the grader made an error. **Send a private Piazza post to only Takahisa Hasegawa and Timothy Bail**. In the Summary add "[Request] Regrade Project 1." In the Details add sufficient explanation as to why you think the grader made a mistake. Be concrete and specific. We will not consider requests that do not follow these directions.

It is important to note that because we consider your ability to internalize feedback a learning goal, we also assess it. This ability is considered 10% of each assignment. We default to assigning you full credit. If you request a regrade and do not receive at least 5 points as a result of the request, you will lose those 10 points.

References

[SB20] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. 2nd Ed. MIT press, 2020. URL: http://incompleteideas.net/book/the-book-2nd.html.

[Sut88] Richard Sutton. "Learning to Predict by the Method of Temporal Differences". In: *Machine Learning* 3 (Aug. 1988), pp. 9–44.