

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.
- Submit your code according to the instructions of the TA.
 - 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 TA.
 - 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 based on the TA's instruction.
- Write a paper **in English** describing the experiments, how you replicated them, and any other relevant information.
 - 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!
 - Finally, under no circumstances should you use code or writing found online, or from another student in the current or a previous semester.

2 Resources

2.1 Lectures

- Lesson 3: Temporal Difference Learning

2.2 Readings

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

3 Submission Details

Due Date: **October 16th, 2024, 23:59 PM**

The submission consists of:

- Your written report in PDF format
- Your source code

3.1 Grading and Regrading

If you are convinced that your grade is in error, 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 explains where the grader made an error.

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