

# REINFORCEMENT LEARNING

---

## FUNDAMENTALS AND APPLICATIONS

### Final Assignment

Topic Due March 1st by 11:59 p.m. EST

Final Submission Due **April 23rd**, by 11:59 p.m. EST

The final assignment involves whichever application of reinforcement learning you find the most personally interesting and motivating. **The topic you choose must be approved by me**, so email me before March 1st with ideas. If you need assistance choosing a topic (e.g., you want to replicate a paper's results but don't know a good paper specifically) I will help.

I will also help you with any aspect of this assignment short of producing submission pieces for you. For example, if you need a lot of help coding part of your assignment I am willing to advise as much as I can. The purpose of the final assignment is for you to come out the other end a more skilled analyst/engineer; we aren't going to Monte Carlo the skills required for this assignment.

Choose one of the following:

1. **Implement the algorithm in a paper of your choice and/or replicate the results.** Submission should include:
  - a. A link to a github repository that contains runnable code for the implementation as well as any necessary data.
  - b. A summary of the paper and its findings.
  - c. A description of your experience with the assignment: obstacles, struggles, frustrations, interesting new learnings, etc.
2. **Create a custom OpenAI Gym Environment for a task of your choice and design, and solve it using one tabular method and one function approximation method** (e.g., tabular Q-Learning and a Q-Network). Submission should include:
  - a. A link to a github repository that contains runnable code for the implementation as well as any necessary data.
  - b. A summary of your task: an overview, the motivation behind your choice, a technical description of the MDP (complete with what you have decided indicates an agent has 'solved' your environment), a technical description of the algorithms you chose to solve your environment, and some visualizations of your algorithms' learning of the environment solution.
  - c. A description of your experience with the assignment: obstacles, struggles, frustrations, interesting new learnings, etc.
3. **An independent project of your choosing that applies RL to a custom area or dataset.** This option is here for students that have an idea that is not covered by the above two options. Some of the students in this course have more experience implementing RL projects and may prefer to add to their chosen area of research/work while getting good grades at the same time.



## Assignment Schedule

Title	Released	Due	Grade %	Notes
Assignment 1: The K-Armed Bandit	January 14th	January 22nd	Written: 1% Programming: 2%	
Assignment 2: Bellman Equations and MDPs	January 22nd	February 5th	Written: 6%	
Assignment 3: Monte Carlo and TD-Learning	February 13th	March 1st	Programming: 4%	
Assignment 4: Advanced Tabular Methods	March 7th	March 22nd	Programming: 4%	
Assignment 5: Function Approximation for Control	March 19th	March 29th	Programming: 5%	
Assignment 6: Policy Gradient Methods	April 9th	April 16th	Programming: 5%	
Final Written Assignment	February 20th	April 23rd	Written: 18%	See PDF on OWL
Final Programming Assignment	February 20th	April 23rd	Programming: 35%	See PDF on OWL.



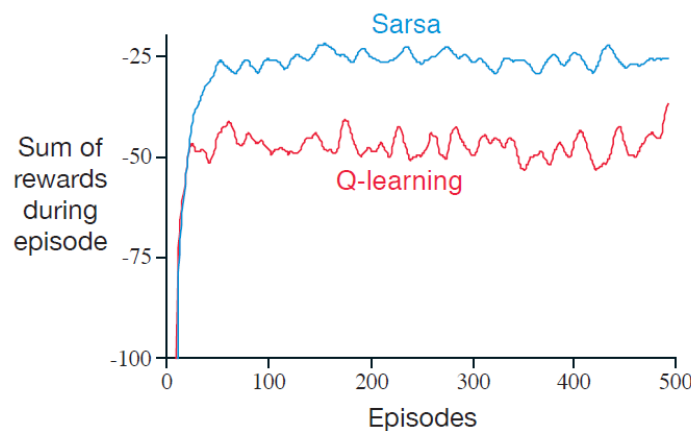
# Assignment 3

Implementation of Monte Carlo and Temporal Difference Learning Algorithms

Due: Monday, March 1st by 11:59pm

For each item, provide a writeup of your findings and your experience. What did you struggle with? What did you learn?

**1a. Recreate the figure below in the text using the 'CliffWalking-v0' task in OpenGym.**



**1b. Add on-policy Monte Carlo. How does it compare?**

**2a. Implement one of the following to learn the Frozen Lake task in OpenGym.** Don't worry if you can't solve it right away. Frozen Lake is a "toy" task, but solving it without changing the reward function is non-trivial.

1. Double Q-Learning
2. Expected SARSA

**2b. What is the reward function for the Frozen Lake task? How does this affect your agent's ability to learn? How might you make things easier for your agent? Modify the reward your agent receives from its actions. Can you achieve better performance?**

- *env.desc* gives you the map. 'H' is a hole, 'G' is the goal, 'S' is the start, and 'F' is the frozen lake.

**2c. Now let's achieve better performance without cheating. Frozen Lake is about exploration. How could you promote better exploration? What's the best performance you think you can get, and why?**

- *You might try a different initialization approach, decay epsilon over time, make your agent wait until to update|decay until it has received a reward of some kind, etc.*

**3. Implement tabular Q-Learning to solve the Mountain Car task in OpenGym.**

- Note that the state is a length 2 vector and is *continuous*. You will have to discretize the state to solve the task. How you discretize the state is up to you.

## Section 1: Student Ratings Summary

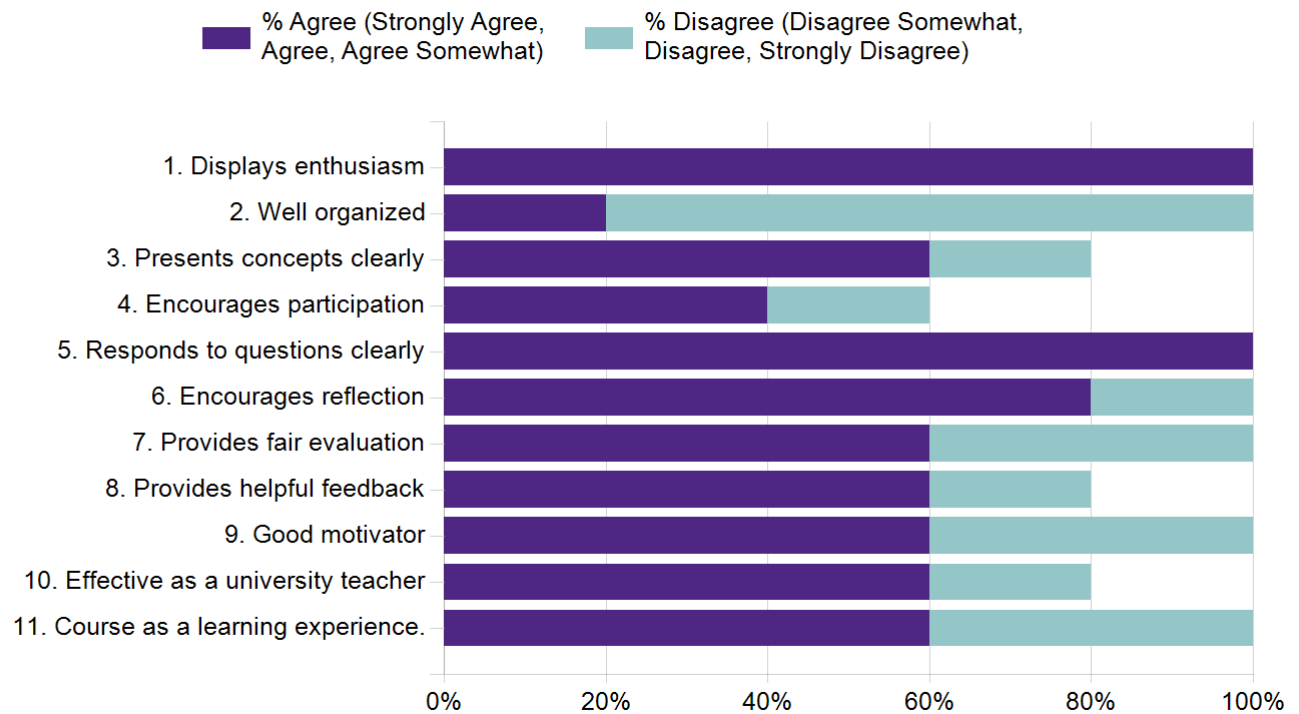
Frequency of Responses for the Core SQCT Questions

	Res p	Strongly Disagree (1)	Disagre e (2)	Disagree Somewhat (3)	Neither Agree nor Disagree (4)	Agree Somewhat (5)	Agree (6)	Strongly Agree (7)
<b>1. Displays enthusiasm</b>	5	0	0	0	0	1	1	3
<b>2. Well organized</b>	5	2	0	2	0	0	0	1
<b>3. Presents concepts clearly</b>	5	1	0	0	1	0	2	1
<b>4. Encourages participation</b>	5	1	0	0	2	0	1	1
<b>5. Responds to questions clearly</b>	5	0	0	0	0	1	2	2
<b>6. Encourages reflection</b>	5	1	0	0	0	1	1	2
<b>7. Provides fair evaluation</b>	5	1	0	1	0	0	1	2
<b>8. Provides helpful feedback</b>	5	0	1	0	1	1	0	2
<b>9. Good motivator</b>	5	2	0	0	0	1	0	2
<b>10. Effective as a university teacher</b>	5	1	0	0	1	0	1	2
<b>11. Course as a learning experience.</b>	5	1	0	1	0	0	1	2

Note. Resp = Total number of responses; Strongly Disagree = 1; Disagree = 2; Disagree Somewhat = 3; Neither Agree nor Disagree = 4, Agree Somewhat = 5; Agree = 6; Strongly Agree = 7.

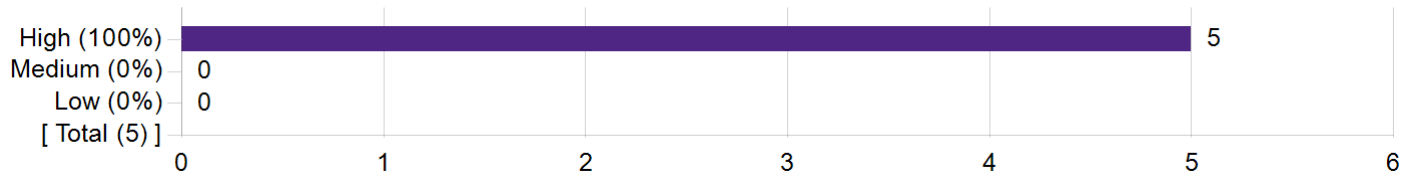
---

## Percentage Agreement and Disagreement for the Core SQCT Questions



## Initial Level of Enthusiasm

Your level of enthusiasm to take this course, at the time of initial registration



## Section 3: Response Demographics

### Reliability Assessment

With 5 responses from a survey population of 11  
the data presented in this report is considered to be **Insufficient**

Number of responses needed to be considered sufficient: 8

Number of responses needed to be considered good: 10

### Student Responses

*Your videos are not well planned in advance  
videoes are too long with many unrelated stories in between which makes it hard to learn the  
main lesson  
assignments are much harder than the course  
midterm was too hard without enough time which caused the problem that you know, and  
created a lot of stress for students*

*Overall the course was a lot of fun. I learnt a lot and I'm certainly glad I took it.*

*Thank you for your effort and help!*

*The course is asynchronous, which is not a good way for the student efficiently to acquire knowledge in RL. I understand that the instructor was trying to make the course video interesting and easy to follow, but it was still difficult for the students to have a solid comprehension of the course materials. I had known some basic concepts of RL before the course, but I think it is not friendly for elementary-level students.*

*The workload in this course was heavy. When I enrolled in this course, it showed "two hours each week". However, the video and the slides made by the instructors didn't provide enough content, the student must spend too much time reading the reference book, and then cause a heavy workload. I felt I just joined a reading group of the reference book. Several weeks after the course's opening, I even gave up watching the course video and slides from the instructors, and just began to learn via the resources on the Internet. This course required too many assignments, which was a heavier workload for us compared to other courses I had taken in Western.*

*However, it is great that the instructor provided some resources, including reading materials and code templates. I think it can help the students easier to practice what they learned in the course.*

*But I believe the instructor can provide better course experiences in the future.*

*This instructor works very hard and spends a lot of time giving us one-on-one instruction. I think he is a very good teacher!*

## Sample Images from Python Notebook

