Course 1, Module 3
Markov Decision Processes

CMPUT 397
Fall 2019

Weekly Schedule

- Sunday: Discussion question & practice quiz
- Monday: Review, Q&A, exercise questions
- Wednesday: In class discussion
- Thursday: Graded Assessment
- Friday: In class exercise questions from worksheet. (finish discussion if needed.)

Reminders: Sept 16, 2019

- Graded Assessment for Course 1, Module 3 (3 MDPs) due **this Thursday**
- **Peer-review for Course 1, Module 3 (3 MDPs) due this Sunday**

Things are getting better :)

- Could you please **explain** question 7 ... ? (Not something you would discuss in a group)
- 7th and 11th question of the quiz (not a question. Request for help...office hours :))
- For the first step of MDP with state S0 and action A0, Why the reward is R1?
- is there other **method** other than **MDP**? (We need more context)
- Can **problems** modelled in **MDPS coverge**?
- When the problem can be defined as both **periodic** and continuing problem, which one is better?
- About the **dynamic programming solution** of Markov decision process: the advantage of dynamic programming is ...

Fast clarifications

- Why the **discount rate** is **less than 1 and grater than 0**?
- Can we use the discount rate method in **episodic tasks** if the number of steps is **extremely large** but finite?
- Can the **possible states be continuous** values instead of a number of **classes**?
- Do we have to use **exponential discounting**?
- If an agent has to choose between two states with both negative long term rewards, **how will the agent know what to choose**?
- What if **state does change** based the agent's **action**?
- Are there **Bandit problems** that **cannot** be represented as **MDPs**?

Let's think on these ...

- How is **K-armed bandit** problem **related** to **mdp**?
- How can we include **subgoals** in our formulation?
- A thermostat is used as an example in the video for Continuing tasks. In this case, if you are using discounted return, **would that mean eventually pressing on thermostat would do nothing** as the gamma term goes to affectively zero?
- It seems **unrealistic** to assume **access to the Markov State**, can we relax this assumption?
- **Doesn't everything end**? How could anything be a continuing task?
- Is it true that the agent can take **all the future reward** into consideration if the discount factor is 1?

Discussion topics for today

1. How do we **choose gamma**? How do we choose the optimal gamma (Xutong, your TA)

2. If there is **noise** or things in the environment are **changing** (e.g., the daily weather) how can we **estimate the reward**? (Ryan, your TA)

3. **What is the Markov property** and what does it mean? (Derek, your TA)

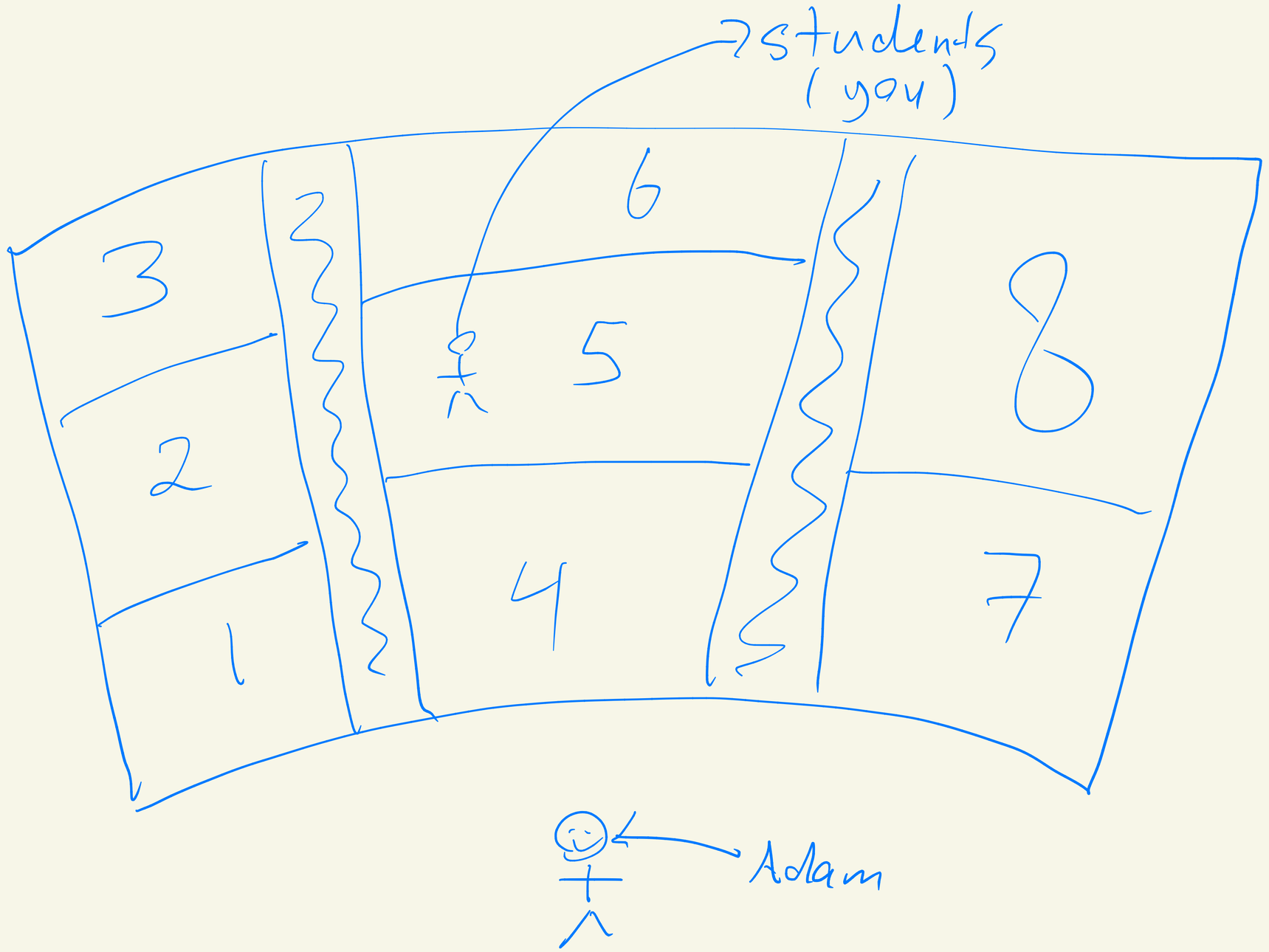
4. How do we decide at **what level to model** the agent and environment interaction? What does that even mean? (Sungsu, your TA)

5. In general there will be **multiple ways** to define the state! How do we know what is a **good state**? How do we decide what is part of the state?

6. Given a problem description how do we decide if the problem is **well modelled as an MDP?**

7. Many real life situations can be modelled as MDPs. Think of some situations where this would not be possible, or **where MDP is just generally not suited to the task**.

8. The MDP's discussed depend on a **human defining the states, actions, and rewards**. Could the **agent** define these **themselves**? How? (Alex, your TA)


7 students
(you)
3
2
6
5
8
2
1
4
7
Adam