4. Utility Function

Utility Function

So what we will get here is our max divided by 1 minus gamma.

And this is a bound.

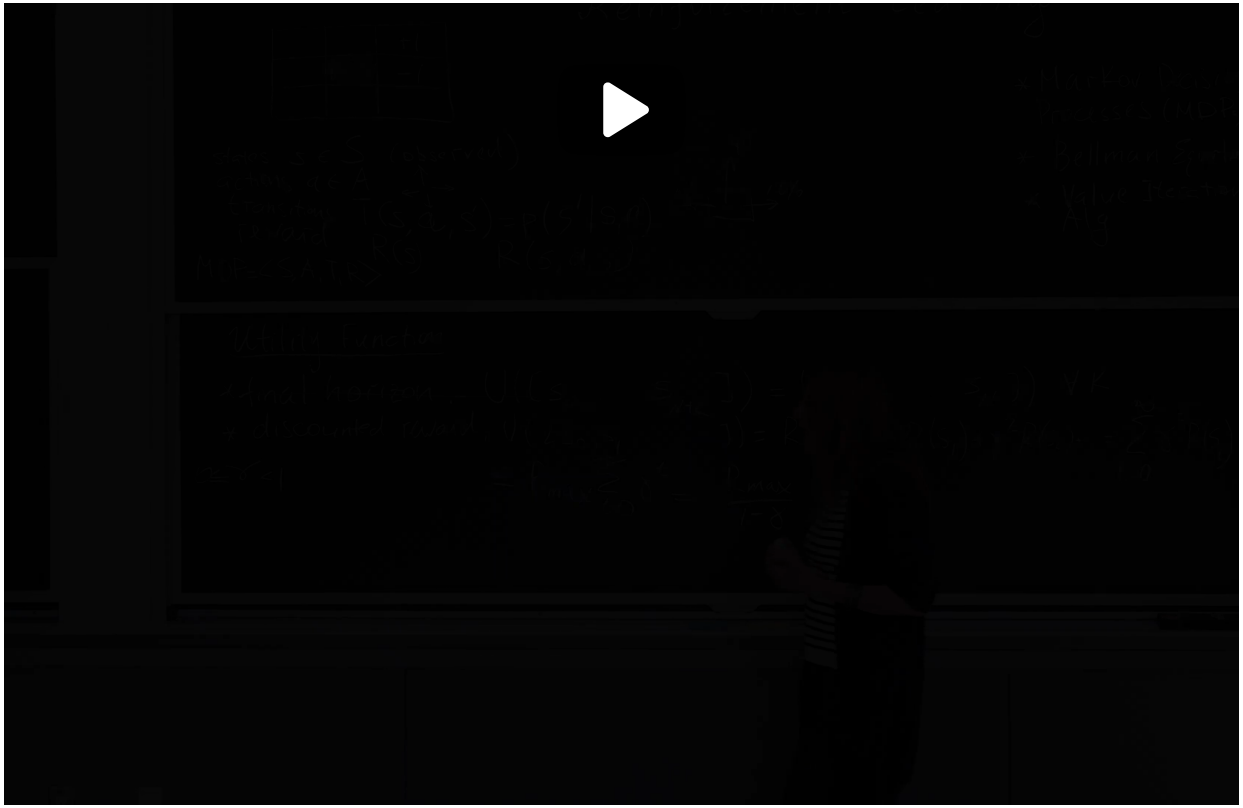
And what you will see is that actually,

the fact that we're using the discount and rewards

would be essential for us to make our algorithms converge.

So this is a discounted reward that we

will be using for the duration.



End of transcript. Skip to the start.

7:43 / 7:43 1.50x

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

Final Horizon vs Discounted Reward

1/1 point (graded)

An RL agent visits states $s_1, s_2 \ldots$ starting from s_0 and receives rewards $r_0, r_1, \ldots$ respectively.

Which of the following options is true about the utility function $U\left([s_1, s_2 \ldots s_n]\right)$

- [x] Final Horizon based utility could be higher by considering different actions from the same state occuring at different time steps ✔
- [ ] Final Horizon based utility cannot be higher by considering different actions from the same state occuring at different time steps
- [x] Discounted Reward based utility cannot be higher by considering different actions from the same state occuring at different time steps ✔
- [ ] Discounted Reward based utility can be higher by considering different actions from the same state occuring at different time steps

✔

Solution:

When using a final horizon utility function, when computing the action at time step i, it can depend on the amount of steps that we have left until we reach N. For example, if we are at state s at time N, the agent will want to act greedily and take the action that leads to the immediate highest reward. However, if we are at time 0, the agent can allow to move towards areas with higher rewards while getting an immediate lower reward.

Discounted Reward based utility under a markovian setting would lead to an optimal policy that only depends on the state and is independent of the step where the state occurs.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Discounted Reward

1/1 point (graded)

Recall that the discounted reward is given by the following formula:

$$R(s_0) + \gamma R(s_1) + \gamma^2 R(s_2) \ldots = \sum_{t=0}^{\infty} \gamma^t R(s_t)$$

Select all the correct statement(s) from option(s) below:

- [x] For $\gamma = 0$, maximizing for discounted reward boils down to greedily maximizing for the immediate reward ✔
- [x] Discounted Reward is guaranteed to be finite when the maximum reward is finite ✔
- [ ] Discounted Reward can be unbounded when the maximum reward is finite

- [ ] Discounted Reward converges to $R_{min} / (1 - \gamma)$ where R_{min} is the minimum reward possible from any state

✔

Solution:

For $\gamma = 0$, the discounted reward is given by,

$$R(s) + 0 * R(s_1) + \ldots = R(s)$$

which is dependent only on the reward for the current step.
When the maximum reward is finite, the discounted reward as derived in the lecture is given by $R_{max} / (1 - \gamma)$

Submit You have used 1 of 3 attempts

ℹ Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 5 Reinforcement Learning (2 weeks) :Lecture 17. Reinforcement Learning 1 / 4. Utility Function

Add a Post

Show all posts ▼ by recent activity ▼

? [Staff] Final Horizon vs Discounted Reward: could you clarify a couple things? 1