CS394R
Reinforcement Learning: Theory and Practice

Scott Niekum and Peter Stone

Department of Computer Science
The University of Texas at Austin
Good Morning Colleagues

- Are there any questions?
Logistics

- Continue towards final project proposal
Logistics

- Continue towards final project proposal
- Almost done with content that will be on midterm
Logistics

- Continue towards final project proposal
- Almost done with content that will be on midterm
- Next week’s readings
Logistics

- Continue towards final project proposal
- Almost done with content that will be on midterm
- Next week’s readings
  - Policy gradient methods
Another way to blend TD $\rightarrow$ MC (other than n-step returns)
Chapter 12 - Eligibility Traces

- Another way to blend $\text{TD} \rightarrow \text{MC}$ (other than n-step returns)
- Equally applicable in continuous and discrete settings
Common Questions

- When do we use online vs offline TD?
- Please discuss true online TD lambda further.
- Please explain the relationship between the forward and backward views.
- Why is TD(\(\lambda\)) an approximation of the off-line \(\lambda\)-return algorithm? Where is the approximation?
Other Common Questions

• How do we set the proper lambda (and other hyperparameters) for TD Learning?
Other Common Questions

• How do we set the proper lambda (and other hyperparameters) for TD Learning?

• Is it possible to combine the $\lambda$ and $\gamma$ parameters?
Other Common Questions

- How do we set the proper lambda (and other hyperparameters) for TD Learning?
- Is it possible to combine the \( \lambda \) and \( \gamma \) parameters?
- Please discuss pseudo-termination further
Other Common Questions

- How do we set the proper lambda (and other hyperparameters) for TD Learning?
- Is it possible to combine the $\lambda$ and $\gamma$ parameters?
- Please discuss pseudo-termination further
  - Predict quantities that aren’t part of the problem
Other Common Questions

• How do we set the proper lambda (and other hyperparameters) for TD Learning?

• Is it possible to combine the $\lambda$ and $\gamma$ parameters?

• Please discuss pseudo-termination further
  – Predict quantities that aren’t part of the problem
  – e.g. reward in 4 steps
  – number of steps to landmark
Other Interesting Questions

- Zhi Wang: Why is it called an *eligibility trace*?
Other Interesting Questions

- Zhi Wang: Why is it called an *eligibility trace*?

- Steve Han: If eligibility traces are superior, why are Q-Learning and TD(0) so widely used??

- Daniel Almeraz: The book mentions eligibility traces are bad in short tasks, and offline tasks with a lot of data. Is there a case where they still end up being the best option?
Other Interesting Questions

- Zhi Wang: Why is it called an *eligibility trace*?

- Steve Han: If eligibility traces are superior, why are Q-Learning and TD(0) so widely used??

- Daniel Almeraz: The book mentions eligibility traces are bad in short tasks, and offline tasks with a lot of data. Is there a case where they still end up being the best option?

- Stephane Hatgis-Kessell: Why are eligibility traces useful for non-Markovian tasks?
Other Interesting Questions

• Zirui Tang: What are the similarities and differences between Monte Carlo methods and TD(1)?
Other Interesting Questions

• Zirui Tang: What are the similarities and differences between Monte Carlo methods and TD(1)?

• Haoqi Wang: How much more expensive (computationally and space) is the online $\lambda$-return algorithm than TD(\(\lambda\))?
Other Interesting Questions

• Zirui Tang: What are the similarities and differences between Monte Carlo methods and TD(1)?

• Haoqi Wang: How much more expensive (computationally and space) is the online $\lambda$-return algorithm than TD($\lambda$)?

• Joseph Muffoletto: When would we want to use variable $\gamma$ or $\lambda$?