

CS394R
Reinforcement Learning:
Theory and Practice

Amy Zhang and Peter Stone

Departments of ECE and CS
The University of Texas at Austin

Good Morning Colleagues

- Are there any (logistics) questions?

Logistics

- Midterm Thursday or Friday - 3 hours timed

Logistics

- Midterm Thursday or Friday - 3 hours timed
- No class Thursday

Logistics

- Midterm Thursday or Friday - 3 hours timed
- No class Thursday
- Feedback on final project proposals coming

Logistics

- Midterm Thursday or Friday - 3 hours timed
- No class Thursday
- Feedback on final project proposals coming
- Next step: literature surveys

Logistics

- Midterm Thursday or Friday - 3 hours timed
- No class Thursday
- Feedback on final project proposals coming
- Next step: literature surveys
 - Build on proposal

Logistics

- Midterm Thursday or Friday - 3 hours timed
- No class Thursday
- Feedback on final project proposals coming
- Next step: literature surveys
 - Build on proposal
- Next week's readings

Logistics

- Midterm Thursday or Friday - 3 hours timed
- No class Thursday
- Feedback on final project proposals coming
- Next step: literature surveys
 - Build on proposal
- Next week's readings
 - Deep RL algorithms, Distributional RL

Logistics

- Midterm Thursday or Friday - 3 hours timed
- No class Thursday
- Feedback on final project proposals coming
- Next step: literature surveys
 - Build on proposal
- Next week's readings
 - Deep RL algorithms, Distributional RL
 - No longer a textbook

Ch.16: Applications and Case Studies

- Many more applications on resources page

Ch.16: Applications and Case Studies

- Many more applications on resources page
- Skipped connections to:
 - Ch.14 psychology
 - Ch.15 neuroscience

Ch.16: Applications and Case Studies

- Many more applications on resources page
- Skipped connections to:
 - Ch.14 psychology
 - Ch.15 neuroscience
- Ch.17 summarizes much of what's to come

Context

- *Applications of reinforcement learning are still far from routine and typically require as much art as science. Making applications easier and more straightforward is one of the goals of current research in reinforcement learning.*

Common Questions

- Doesn't DQN meet conditions of deadly triad?

Common Questions

- Doesn't DQN meet conditions of deadly triad?
 - Yes!

Common Questions

- Doesn't DQN meet conditions of deadly triad?
 - Yes!
 - So how is it stable?

Common Questions

- Doesn't DQN meet conditions of deadly triad?
 - Yes!
 - So how is it stable?
- Clips TD-error to $(-1,1)$

Common Questions

- Doesn't DQN meet conditions of deadly triad?
 - Yes!
 - So how is it stable?
- Clips TD-error to $(-1,1)$
- What is *experience replay*? Why use it?

Common Questions

- Doesn't DQN meet conditions of deadly triad?
 - Yes!
 - So how is it stable?
- Clips TD-error to $(-1, 1)$
- What is *experience replay*? Why use it?
 - like Dyna
 - allows the samples not to be strongly correlated

Common Questions

- Doesn't DQN meet conditions of deadly triad?
 - Yes!
 - So how is it stable?
- Clips TD-error to $(-1, 1)$
- What is *experience replay*? Why use it?
 - like Dyna
 - allows the samples not to be strongly correlated
- DQN: How does using a target network help?

Common Questions

- Doesn't DQN meet conditions of deadly triad?
 - Yes!
 - So how is it stable?
- Clips TD-error to $(-1, 1)$
- What is *experience replay*? Why use it?
 - like Dyna
 - allows the samples not to be strongly correlated
- DQN: How does using a target network help?
 - Avoids chasing a moving target