

CS394R

Reinforcement Learning: Theory and Practice

Peter Stone

Department of Computer Science
The University of Texas at Austin

Good Morning Colleagues

- Are there any questions?

Logistics

- Do programming assignments!

Logistics

- Do programming assignments!
- Next week's readings

Logistics

- Do programming assignments!
- Next week's readings
 - **Approximate** on-policy prediction and control

SARSA vs. Q

- Week 0 example
 - (Remember no access to real model)
 - $\alpha = .1$, ϵ -greedy $\epsilon = .75$, break ties in favor of $\rightarrow$

SARSA vs. Q

- Week 0 example
 - (Remember no access to real model)
 - $\alpha = .1$, ϵ -greedy $\epsilon = .75$, break ties in favor of $\rightarrow$
 - Where did policy change?

SARSA vs. Q

- Week 0 example
 - (Remember no access to real model)
 - $\alpha = .1$, ϵ -greedy $\epsilon = .75$, break ties in favor of $\rightarrow$
 - Where did policy change?
- How do their convergence guarantees differ?

SARSA vs. Q

- Week 0 example
 - (Remember no access to real model)
 - $\alpha = .1$, ϵ -greedy $\epsilon = .75$, break ties in favor of $\rightarrow$
 - Where did policy change?
- How do their convergence guarantees differ?
 - Sarsa depends on policy's dependence on Q:
 - Policy must converge to greedy

SARSA vs. Q

- Week 0 example
 - (Remember no access to real model)
 - $\alpha = .1$, ϵ -greedy $\epsilon = .75$, break ties in favor of $\rightarrow$
 - Where did policy change?
- How do their convergence guarantees differ?
 - Sarsa depends on policy's dependence on Q :
 - Policy must converge to greedy
 - Q-learning value function converges to Q^*
 - As long as all state-action pairs visited infinitely
 - And step-size satisfies stochastic convergence equations

More SARSA vs. Q

- Why does Q-learning learn to hug the cliff? (p. 139)

More SARSA vs. Q

- Why does Q-learning learn to hug the cliff? (p. 139)
- Why can expected SARSA use high α , but SARSA can't? (p. 140)