CS395T
Reinforcement Learning: Theory and Practice
Fall 2004

Peter Stone

Department or Computer Sciences
The University of Texas at Austin

Week4a: Tuesday, September 21st

Good Afternoon Colleagues

- Are there any questions?

Some more on Chapter 3

- Consider the week 0 environment

- For some s, what is $V(s)$?

Some more on Chapter 3

- Consider the week 0 environment
- For some s, what is $V(s)$?
- OK - consider the policy we ended with
- Now, for some s, what is $V(s)$?

Some more on Chapter 3

- Consider the week 0 environment
- For some s, what is $V(s)$?
- OK - consider the policy we ended with
- Now, for some s, what is $V(s)$?
- Construct V in undiscounted, episodic case

Some more on Chapter 3

- Consider the week 0 environment
- For some s, what is $V(s)$?
- OK - consider the policy we ended with
- Now, for some s, what is $V(s)$?
- Construct V in undiscounted, episodic case
- Construct Q in undiscounted, episodic case

Some more on Chapter 3

- Consider the week 0 environment
- For some s, what is $V(s)$?
- OK - consider the policy we ended with
- Now, for some s, what is $V(s)$?
- Construct V in undiscounted, episodic case
- Construct Q in undiscounted, episodic case
- What if it's discounted?

Some more on Chapter 3

- Consider the week 0 environment
- For some s, what is $V(s)$?
- OK - consider the policy we ended with
- Now, for some s, what is $V(s)$?
- Construct V in undiscounted, episodic case
- Construct Q in undiscounted, episodic case
- What if it's discounted?
- What if it's continuing?

Value Functions

- Relationship between Q and V

Value Functions

- Relationship between Q and V
- Bellman equations

Value Functions

- Relationship between Q and V
- Bellman equations — unique solution

Value Functions

- Relationship between Q and V
- Bellman equations — unique solution
- Backup diagrams (p. 70, 74, 77)

Value Functions

- Relationship between Q and V
- Bellman equations — unique solution
- Backup diagrams (p. 70, 74, 77)
- Exercise 3.17

This Chapter

- Solution methods **given a model**

This Chapter

- Solution methods **given a model**
- Why is it called dynamic programming?

Student Discussion

- Susan on the Gambler's Problem (p. 101)

Student Discussion

- Susan on the Gambler's Problem (p. 101)
- Email discussion linked to the book web page

Existence/Convergence guarantees

- V^π exists and is unique if $\gamma < 1$ or termination guaranteed for all states under policy π. (p. 90)

Existence/Convergence guarantees

- V^π exists and is unique if $\gamma < 1$ or termination guaranteed for all states under policy π. (p. 90)

- Policy evaluation converges under the same conditions (p. 91)

Existence/Convergence guarantees

- V^π exists and is unique if $\gamma < 1$ or termination guaranteed for all states under policy π. (p. 90)
- Policy evaluation converges under the same conditions (p. 91)
- Exercises 4.1, 4.2

Existence/Convergence guarantees

- V^π exists and is unique if $\gamma < 1$ or termination guaranteed for all states under policy π. (p. 90)
- Policy evaluation converges under the same conditions (p. 91)
- Exercises 4.1, 4.2
- Policy improvement theorem:
$\forall s, Q^\pi(s, \pi'(s)) \geq V^\pi(s) \Rightarrow \forall s, V^{\pi'}(s) \geq V^\pi(s)$

Some more points

- p. 107: Is LP still inefficient?

Some more points

- p. 107: Is LP still inefficient?
- p. 109: This chapter treats **bootstrapping** with a model

Some more points

- p. 107: Is LP still inefficient?
- p. 109: This chapter treats **bootstrapping** with a model
  - Next: no model and no bootstrapping

Some more points

- p. 107: Is LP still inefficient?
- p. 109: This chapter treats **bootstrapping** with a model
  - Next: no model and no bootstrapping
  - Then: no model, but bootstrapping