CS394R
Reinforcement Learning: Theory and Practice
Fall 2007

Peter Stone

Department of Computer Sciences
The University of Texas at Austin

Good Afternoon Colleagues

Good Afternoon Colleagues

- Are there any questions?

Logistics

- Any lingering questions from last week?

This Chapter

- Defines the problem

This Chapter

- Defines the problem
- Introduces some important notation and concepts.

This Chapter

- Defines the problem
- Introduces some important notation and concepts.
  - Returns
  - Markov property
  - State/action value functions
  - Bellman equations

This Chapter

- Defines the problem
- Introduces some important notation and concepts.
  - Returns
  - Markov property
  - State/action value functions
  - Bellman equations
  - Get comfortable with them!

This Chapter

- Defines the problem
- Introduces some important notation and concepts.
  - Returns
  - Markov property
  - State/action value functions
  - Bellman equations
  - Get comfortable with them!
- Solution methods come next

This Chapter

- Defines the problem
- Introduces some important notation and concepts.
  - Returns
  - Markov property
  - State/action value functions
  - Bellman equations
  - Get comfortable with them!
- Solution methods come next
  - What does it mean to **solve** an RL problem?

Formulating the RL problem

- Art more than science

Formulating the RL problem

- Art more than science
- States, actions, rewards

Formulating the RL problem

- Art more than science
- States, actions, rewards
- Rewards: no hints on **how** to solve the problem

Formulating the RL problem

- Art more than science
- States, actions, rewards
- Rewards: no hints on **how** to solve the problem
  - Dependent on next state (p. 66)

Formulating the RL problem

- Art more than science
- States, actions, rewards
- Rewards: no hints on **how** to solve the problem
  - Dependent on next state (p. 66)
- Discounted vs. non-discounted

Formulating the RL problem

- Art more than science
- States, actions, rewards
- Rewards: no hints on **how** to solve the problem
  - Dependent on next state (p. 66)
- Discounted vs. non-discounted
- Episodic vs. continuing

Formulating the RL problem

- Art more than science
- States, actions, rewards
- Rewards: no hints on **how** to solve the problem
  - Dependent on next state (p. 66)
- Discounted vs. non-discounted
- Episodic vs. continuing
- Exercises 3.4, 3.5 (p.59)

Markov property

- What is it?

Markov property

- What is it?
- Does it hold in the real world?

Markov property

- What is it?
- Does it hold in the real world?
- It's an ideal
  - Will allow us to prove properties of algorithms
  - Algorithms may still work when not provably correct

Markov property

- What is it?
- Does it hold in the real world?
- It's an ideal
  - Will allow us to prove properties of algorithms
  - Algorithms may still work when not provably correct
- Exercise 3.6

Value functions

- Consider the week 0 environment

Value functions

- Consider the week 0 environment
- For some s, what is $V(s)$?

Value functions

- 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 $V(s)$?

Value functions

- 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

Value functions

- 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

Value functions

- 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?

Value functions

- 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

- 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?
- Exercises 3.10, 3.11, 3.17