CS395T
Reinforcement Learning: Theory and Practice
Fall 2004

Peter Stone

Department or Computer Sciences
The University of Texas at Austin

Week5b: Thursday, September 30th

Good Afternoon Colleagues

- Are there any questions?

Good Afternoon Colleagues

- Are there any questions?
- Pending questions:
  - Blackjack: why not DP?
  - Figure 5.2: why flat horizontally?
  - Does MC require Markov assumption?
  - Off-policy learning: why just from the tails?
  - What's up with those equations after (5.3)?

Advertisements

- Pazzani talk at 2pm today
- Teller talk at 11am tomorrow

Relationship to DP

- MC doesn't need a (full) model
  - Can learn from actual or simulated experience

Relationship to DP

- MC doesn't need a (full) model
  - Can learn from actual or simulated experience
- DP takes advantage of a full model
  - Doesn't need **any** experience

Relationship to DP

- MC doesn't need a (full) model
  - Can learn from actual or simulated experience
- DP takes advantage of a full model
  - Doesn't need **any** experience
- MC expense independent of number of states

Relationship to DP

- MC doesn't need a (full) model
  - Can learn from actual or simulated experience
- DP takes advantage of a full model
  - Doesn't need **any** experience
- MC expense independent of number of states
- No bootstrapping in MC

Relationship to DP

- MC doesn't need a (full) model
  - Can learn from actual or simulated experience
- DP takes advantage of a full model
  - Doesn't need **any** experience
- MC expense independent of number of states
- No bootstrapping in MC
  - Not harmed by Markov violations

Blackjack

- Fig. 5.2 (114): Why values mainly independent of dealer showing?

Blackjack

- Fig. 5.2 (114): Why values mainly independent of dealer showing?
- As true in Fig. 5.5? (121)

Blackjack

- Fig. 5.2 (114): Why values mainly independent of dealer showing?
- As true in Fig. 5.5? (121)
- Possible explanation for notch in usable ace policy?

Blackjack

- Fig. 5.2 (114): Why values mainly independent of dealer showing?
- As true in Fig. 5.5? (121)
- Possible explanation for notch in usable ace policy?
- Why not just use DP?

Control

- Q more useful than V without a model

Control

- Q more useful than V without a model
- But to get it need to explore

Control

- Q more useful than V without a model
- But to get it need to explore
- Exploring starts vs. stochastic policies

Control

- Q more useful than V without a model
- But to get it need to explore
- Exploring starts vs. stochastic policies
  - Does ES converge?

Control

- Q more useful than V without a model
- But to get it need to explore
- Exploring starts vs. stochastic policies
  – Does ES converge?(Tsitsiklis paper)

Control

- Q more useful than V without a model
- But to get it need to explore
- Exploring starts vs. stochastic policies
  - Does ES converge?(Tsitsiklis paper)
  - Epsilon-soft vs. epsilon-greedy (122)

Control

- Q more useful than V without a model
- But to get it need to explore
- Exploring starts vs. stochastic policies
  - Does ES converge?(Tsitsiklis paper)
  - Epsilon-soft vs. epsilon-greedy (122)
  - Why consider off-policy methods?

Learning off policy

- Off policy equations (5.3 and next 2: 125)

Learning off policy

- Off policy equations (5.3 and next 2: 125)
- Why only learn from tail in Fig. 5.7?

Learning off policy

- Off policy equations (5.3 and next 2: 125)
- Why only learn from tail in Fig. 5.7?
- Off policy possible with policy search (evolution)?