CS394R
Reinforcement Learning: Theory and Practice

Peter Stone

Department of Computer Science
The University of Texas at Austin

Good Afternoon Colleagues

- Are there any questions?

Good Afternoon Colleagues

- Are there any questions?

Logistics

- Next week - multiagent learning

Logistics

- Next week - multiagent learning
- Final projects due next Thursday at the beginning of class, but...

Logistics

- Next week - multiagent learning
- Final projects due next Thursday at the beginning of class, but...
- 4pm Friday is when I need them to be able to look on the way home

Logistics

- Next week - multiagent learning
- Final projects due next Thursday at the beginning of class, but...
- 4pm Friday is when I need them to be able to look on the way home
- So *if in class on Thursday*, due electronically Friday at 4pm

Logistics

- Next week - multiagent learning
- Final projects due next Thursday at the beginning of class, but...
- 4pm Friday is when I need them to be able to look on the way home
- So *if in class on Thursday*, due electronically Friday at 4pm
- Also put one hard copy outside my office by then.

Logistics

- Next week - multiagent learning
- Final projects due next Thursday at the beginning of class, but...
- 4pm Friday is when I need them to be able to look on the way home
- So *if in class on Thursday*, due electronically Friday at 4pm
- Also put one hard copy outside my office by then.
- After that, considered late.

Least Squares Methods

- On-policy Least squares projection could cause Policy Iteration to diverge.

Least Squares Methods

- On-policy Least squares projection could cause Policy Iteration to diverge.
- Example from Koller and Parr 2000:

Least Squares Methods

- On-policy Least squares projection could cause Policy Iteration to diverge.
- Example from Koller and Parr 2000:
  - 4 states: s_0, s_1, s_2, s_3
  - 2 actions: R, L (10% chance of moving opposite directions)
  - Rewards of +1 in states s_1 and s_2
  - basis functions are 1, x, x^2 (mapping from s_x)

Least Squares Methods

- On-policy Least squares projection could cause Policy Iteration to diverge.
- Example from Koller and Parr 2000:
  - 4 states: s_0, s_1, s_2, s_3
  - 2 actions: R, L (10% chance of moving opposite directions)
  - Rewards of +1 in states s_1 and s_2
  - basis functions are 1, x, x^2 (mapping from s_x)
  - Possible functions are parabolas

Least Squares Methods

- On-policy Least squares projection could cause Policy Iteration to diverge.
- Example from Koller and Parr 2000:
  - 4 states: s_0, s_1, s_2, s_3
  - 2 actions: R, L (10% chance of moving opposite directions)
  - Rewards of +1 in states s_1 and s_2
  - basis functions are 1, x, x^2 (mapping from s_x)
  - Possible functions are parabolas
  - Starting with RRRR policy leads to poor approximation (graph in paper), iterates to LLLL and oscillates

Least Squares Methods

- On-policy Least squares projection could cause Policy Iteration to diverge.
- Example from Koller and Parr 2000:
  - 4 states: s_0, s_1, s_2, s_3
  - 2 actions: R, L (10% chance of moving opposite directions)
  - Rewards of +1 in states s_1 and s_2
  - basis functions are 1, x, x^2 (mapping from s_x)
  - Possible functions are parabolas
  - Starting with RRRR policy leads to poor approximation (graph in paper), iterates to LLLL and oscillates
  - Mainly because stationary distribution rarely visits states s_0 and s_1.