Good Afternoon Colleagues

- Are there any questions?
Good Afternoon Colleagues

- Are there any questions?
- Pending questions:
  - Policy iteration vs. explore/exploit?
Good Afternoon Colleagues

• Are there any questions?

• Pending questions:
  – Policy iteration vs. explore/exploit?
  – Jack’s Car rental pictures
Good Afternoon Colleagues

- Are there any questions?
- Pending questions:
  - Policy iteration vs. explore/exploit?
  - Jack’s Car rental pictures
  - Convergence guarantees (polynomial)
Policy Evaluation

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

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

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

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

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

- Policy evaluation on the week 0 problem
  - Are the conditions met?
Policy Evaluation

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

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

• Policy evaluation on the week 0 problem
  – Are the conditions met?
  – (book slides)
Policy Evaluation

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

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

• Policy evaluation on the week 0 problem
  – Are the conditions met?
  – (book slides)

• Exercises 4.1, 4.2
Policy Improvement

- Policy improvement theorem:
  \[ \forall s, Q_\pi(s, \pi'(s)) \geq V_\pi(s) \Rightarrow \forall s, V_\pi'(s) \geq V_\pi(s) \]
Policy Improvement

• Policy improvement theorem:
  \[ \forall s, Q^\pi(s, \pi'(s)) \geq V^\pi(s) \Rightarrow \forall s, V^\pi'(s) \geq V^\pi(s) \]

• (book slides)
Policy Improvement

- Policy improvement theorem:
  \[ \forall s, Q^\pi(s, \pi'(s)) \geq V^\pi(s) \Rightarrow \forall s, V^{\pi'}(s) \geq V^\pi(s) \]

- (book slides)

- Polynomial time convergence (in number of states and actions) even though \( m^n \) policies.
  - Ignoring effect of \( \gamma \) and bits to represent rewards/transition...
Policy Improvement

- Policy improvement theorem:
  \[ \forall s, Q^\pi(s, \pi'(s)) \geq V^\pi(s) \Rightarrow \forall s, V^\pi'(s) \geq V^\pi(s) \]

- (book slides)

- Polynomial time convergence (in number of states and actions) even though \( m^n \) policies.
  - Ignoring effect of \( \gamma \) and bits to represent rewards/transitions
  - p. 107: Is LP still inefficient?
Value Iteration on Week 0 problem

- Show the new policy at each step
  - Not actually to compute policy
Value Iteration on Week 0 problem

- Show the new policy at each step
  - Not actually to compute policy
  - Break policy ties with equiprobable actions
Value Iteration on Week 0 problem

- Show the new policy at each step
  - Not actually to compute policy
  - Break policy ties with equiprobable actions
  - No stochastic transitions
Value Iteration on Week 0 problem

- Show the new policy at each step
  - Not actually to compute policy
  - Break policy ties with equiprobable actions
  - No stochastic transitions

- What happens if we output deterministic policy (as in book)?
Value Iteration on Week 0 problem

- Show the new policy at each step
  - Not actually to compute policy
  - Break policy ties with equiprobable actions
  - No stochastic transitions

- What happens if we output deterministic policy (as in book)?

- How would policy iteration proceed in comparison?
  - More or fewer policy updates?
Value Iteration on Week 0 problem

- Show the new policy at each step
  - Not actually to compute policy
  - Break policy ties with equiprobable actions
  - No stochastic transitions

- What happens if we output deterministic policy (as in book)?

- How would policy iteration proceed in comparison?
  - More or fewer policy updates?
  - True in general?
Summary

• p. 109: This chapter treats *bootstrapping* with a model
Summary

• p. 109: This chapter treats **bootstrapping** with a model
  – Next: no model and no bootstrapping
Summary

- p. 109: This chapter treats **bootstrapping** with a model
  - Next: no model and no bootstrapping
  - Then: no model, but bootstrapping