CS 182
Reinforcement Learning
An example RL domain

• Solitaire
  – What is the state space?
  – What are the actions?
  – What is the transition function?
    • Is it deterministic?
  – What are the rewards?
• (What about Tetris?)
MDPs

- Markov Decision Processes
- What makes them "Markov"?
- General routine
  - Start with a state, s
  - $a = \pi(s)$
  - $s' = T(s,a)$
  - $r = R(s,a,s')$
  - $s = s'$; repeat
Policies and values

• What are policies?
• What are value functions?
• How are they related?
Bellman equation

• How are $V(s)$ and $Q(s,a)$ related?
Reward and utility

- Do you keep track of utility?
- Do you have a value function $V(s)$ or $Q(s,a)$?
- How do you value future rewards?
Policies etc.

- Consider “micro pac-man world”
  - 4 squares, 1 ghost, move in 4 cardinal directions or stay still
  - What's a reasonable policy for the domain?
  - What are the Q-values for this policy?
  - What would the RL algorithms do from here?
    - value iteration a.k.a. dynamic programming
    - Q-learning
Issues with RL

• What happens when the state space gets big?
  – or continuous?
• What if there's someone else in the environment?
• How do you learn faster than thousands of trials?