10 worst mistakes of 1st time job hunters (from WSJ)

1. “I would have started looking for jobs earlier”
2. “I would have actually networked”
3. “I would have taken on a job or an internship in addition to my course load”
4. “I would have gotten more involved in career-relevant extracurricular activities”
5. “I would have applied to more jobs”
6. “I would have focused more on becoming ‘professional’”
7. “I would have done more to figure out what my career goals were”
8. “I would have gone to the career center”
9. “I would have kept better track of my achievements”
10. “I would have focused more on developing relevant skills”

Q-Learning

- Q-Learning: sample-based Q-value iteration
- Learn Q*(s,a) values
  - Receive a sample (s,a,s',r)
  - Consider your old estimate: $Q(s,a)$
  - Consider your new sample estimate:

$$Q^*(s,a) = \sum_{s'} T(s,a,s') \left[R(s,a,s') + \gamma \max_{a'} Q^*(s',a') \right]$$

$$sample = R(s,a,s') + \gamma \max_{a'} Q(s',a')$$

  - Incorporate the new estimate into a running average:

$$Q(s,a) \leftarrow (1-\alpha)Q(s,a) + (\alpha)\,[sample]$$

Exploration / Exploitation

- Several schemes for forcing exploration
  - Simplest: random actions (ε greedy)
    - Every time step, flip a coin
    - With probability ε, act randomly
    - With probability 1-ε, act according to current policy
  - Problems with random actions?
    - You do explore the space, but keep thrashing around once learning is done
    - One solution: lower ε over time
    - Another solution: exploration functions

Q-Learning

- Q-learning produces tables of q-values:

http://ccl.northwestern.edu/netlogo/models/community/Reinforcement%20Learning%20Maze

The Story So Far: MDPs and RL

Things we know how to do:

- If we know the MDP
  - Compute V*, Q*, π* exactly
  - Evaluate a fixed policy π
- If we don't know the MDP
  - We can estimate the MDP then solve
  - We can estimate V for a fixed policy π
  - We can estimate Q*(s,a) for the optimal policy while executing an exploration policy

Techniques:

- Model-based DPs
  - Value and policy Iteration
  - Policy evaluation
- Model-based RL
- Model-free RL:
  - Value learning
  - Q-learning

Midterm

- Questions help to give sense of what's important?
  - Anything in reading is fair game
  - But focus on what we emphasized in class & in projects