CS394R
Reinforcement Learning: Theory and Practice

Amy Zhang and Peter Stone

Departments of ECE and CS
The University of Texas at Austin

Good Morning Colleagues

Good Morning Colleagues

- Are there any questions?

Logistics

Logistics

- Registering for the course

Logistics

- Registering for the course
- If you missed Tuesday . . .

Logistics

- Registering for the course
- If you missed Tuesday . . .
  - Watch intro lecture video
  - Read webpage carefully
- Don't fall behind!

Logistics

- Registering for the course
- If you missed Tuesday . . .
  - Watch intro lecture video
  - Read webpage carefully
- Don't fall behind!
- Email both instructors and TAs

Logistics

- Registering for the course
- If you missed Tuesday . . .
  - Watch intro lecture video
  - Read webpage carefully
- Don't fall behind!
- Email both instructors and TAs
- Class participation:

Logistics

- Registering for the course
- If you missed Tuesday . . .
  - Watch intro lecture video
  - Read webpage carefully
- Don't fall behind!
- Email both instructors and TAs
- Class participation:
  - Attendance (no excuses, but no penalty for missing 3 classes)

Logistics

- Registering for the course
- If you missed Tuesday . . .
    - Watch intro lecture video
    - Read webpage carefully
- Don't fall behind!
- Email both instructors and TAs
- Class participation:
    - Attendance (no excuses, but no penalty for missing 3 classes)
    - Office hours, discussion boards, active in class (name tents coming)

Logistics

- Registering for the course
- If you missed Tuesday . . .
    - Watch intro lecture video
    - Read webpage carefully
- Don't fall behind!
- Email both instructors and TAs
- Class participation:
    - Attendance (no excuses, but no penalty for missing 3 classes)
    - Office hours, discussion boards, active in class (name tents coming)
    - Will seek ways to *give* credit, not dock

More Logistics

More Logistics

- Next readings:

More Logistics

- Next readings:

    – **2nd edition!!**

More Logistics

- Next readings:

  - **2nd edition!!**
  - MDPs and Dynamic Programming

More Logistics

- Next readings:

  - **2nd edition!!**
  - MDPs and Dynamic Programming
  - Budget a good amount of time!

More Logistics

- Next readings:

  - **2nd edition!!**
  - MDPs and Dynamic Programming
  - Budget a good amount of time!
  - Mostly chapter 3 Tuesday, then chapter 4

More Logistics

- Next readings:

  - **2nd edition!!**
  - MDPs and Dynamic Programming
  - Budget a good amount of time!
  - Mostly chapter 3 Tuesday, then chapter 4
  - Single written response to cover both.

More Logistics

- Next readings:

  - **2nd edition!!**
  - MDPs and Dynamic Programming
  - Budget a good amount of time!
  - Mostly chapter 3 Tuesday, then chapter 4
  - Single written response to cover both.

- Do the first exercises and programming assignment

More Logistics

- Next readings:

  - **2nd edition!!**
  - MDPs and Dynamic Programming
  - Budget a good amount of time!
  - Mostly chapter 3 Tuesday, then chapter 4
  - Single written response to cover both.

- Do the first exercises and programming assignment

- Post an intro on the discussion board

More Logistics

- Next readings:

  - **2nd edition!!**
  - MDPs and Dynamic Programming
  - Budget a good amount of time!
  - Mostly chapter 3 Tuesday, then chapter 4
  - Single written response to cover both.

- Do the first exercises and programming assignment

- Post an intro on the discussion board

- Look at resources page

Reading Responses

- Nice responses!

Reading Responses

- Nice responses!
    - Be clear and specific

Reading Responses

- Nice responses!
    - Be clear and specific
    - Short and focussed is fine

Reading Responses

- Nice responses!

  - Be clear and specific
  - Short and focussed is fine
  - Develop your own style

Reading Responses

- Nice responses!

    - Be clear and specific
    - Short and focussed is fine
    - Develop your own style
    - Help us help you

Reading Responses

- Nice responses!

  - Be clear and specific
  - Short and focussed is fine
  - Develop your own style
  - Help us help you
  - Also ask in class or on discussion board

Our Role

Our Role

- Our role isn't to teach RL

Our Role

- Our role isn't to teach RL

- It's to help **you** learn RL

Our Role

- Our role isn't to teach RL

- It's to help **you** learn RL

  – provide context
  – guide your learning (assign readings, exercises, activities)
  – clarify misconceptions

Our Role

- Our role isn't to teach RL

- It's to help **you** learn RL

  - provide context
  - guide your learning (assign readings, exercises, activities)
  - clarify misconceptions

- You have to do the learning

Our Role

- Our role isn't to teach RL

- It's to help **you** learn RL

  - provide context
  - guide your learning (assign readings, exercises, activities)
  - clarify misconceptions

- You have to do the learning

- Read, write, ask, answer, program (investigate)

Let's Play!

Let's Play!

- I'm a 2-armed bandit

Let's Play!

- I'm a 2-armed bandit
- As a class, you choose which arm

Let's Play!

- I'm a 2-armed bandit
- As a class, you choose which arm
- Maximize your payoff.

Let's Play!

- I'm a 2-armed bandit
- As a class, you choose which arm
- Maximize your payoff.
- The answer:

Let's Play!

- I'm a 2-armed bandit
- As a class, you choose which arm
- Maximize your payoff.
- The answer:

```
(defun l () (+ 5 (random 7)))          expectation: 8


(defun r ()
   (let ((x (random 4)))
     (case x
       (0 20) (1 0) (2 0)
       (3 (+ 7 (random 11))))))))       expectation: 8.5
```

Let's Play!

- I'm a 2-armed bandit
- As a class, you choose which arm
- Maximize your payoff.
- The answer:

```
(defun l () (+ 5 (random 7)))        expectation: 8


(defun r ()
   (let ((x (random 4)))
     (case x
        (0 20) (1 0) (2 0)
        (3 (+ 7 (random 11))))))) )    expectation: 8.5
```

- What about minimizing risk?

N-armed bandit in practice?

N-armed bandit in practice?

- Choosing mechanics

- Choosing a barber/hairdresser

N-armed bandit in practice?

- Choosing mechanics

- Choosing a barber/hairdresser

stationary or non-stationary?

Shivaram's Slides

Shivaram's Slides

- Why are they called "bandit" algorithms?

- What changes if we dont have infinite attempts at the bandits, but a limited amount. How should we weight exploration vs exploitation then?

Gradient Bandits

Assignments

- Monitor and contribute to discussion forums!

Assignments

- Monitor and contribute to discussion forums!

- 1st exercises and programming assignment

Assignments

- Monitor and contribute to discussion forums!

- 1st exercises and programming assignment

- Read Chapters 3 and 4

Assignments

- Monitor and contribute to discussion forums!

- 1st exercises and programming assignment

- Read Chapters 3 and 4

- Submit a reading response by 5pm Monday