CS325 Artificial Intelligence
Markov Decision Process Review

Cengiz Günay, Emory Univ.

(AN UNMATCHED LEFT PARENTHESIS CREATES AN UNRESOLVED TENSION THAT WILL STAY WITH YOU ALL DAY.

Spring 2013

CajunBot I and II:

Markov Decision Processes (MDPs) Exercises

$$V(s) \leftarrow \left[\arg\max_a \gamma \sum_{s'} P(s'|a) V(s') \right] + R(s).$$

	1	2	3	4
a				
b	-100			$+100$

$$V(s) \leftarrow \left[\arg\max_a \gamma \sum_{s'} P(s'|a)V(s')\right] + R(s).$$

	1	2	3	4
a				
b	-100			$+100$

Movement probability:

- p succeeds
- $(1-p)$ fails, go reverse

$$V(s) \leftarrow \left[\arg\max_a \gamma \sum_{s'} P(s'|a)V(s')\right] + R(s).$$

	1	2	3	4
a				
b	-100			$+100$

Movement probability:

- p succeeds
- $(1 - p)$ fails, go reverse

Question: **Find final values of all positions**.

-

$$V(s) \leftarrow \left[\arg \max_a \gamma \sum_{s'} P(s'|a) V(s') \right] + R(s) \,.$$

	1	2	3	4
a				
b	-100			$+100$

Movement probability:

- p succeeds
- $(1 - p)$ fails, go reverse

Question: **Find final values of all positions**.

- For $p = 1$, $\gamma = 1$, $\text{cost} = -4$

$$V(s) \leftarrow \left[\arg\max_a \gamma \sum_{s'} P(s'|a)V(s')\right] + R(s)\,.$$

	1	2	3	4
a	84	88	92	96
b	-100	92	96	$+100$

Movement probability:

- p succeeds
- $(1-p)$ fails, go reverse

Question: **Find final values of all positions**.

- For $p = 1$, $\gamma = 1$, $\text{cost} = -4$

$$V(s) \leftarrow \left[\arg\max_a \gamma \sum_{s'} P(s'|a)V(s')\right] + R(s)\,.$$

	1	2	3	4
a				
b	-100			$+100$

Movement probability:

- p succeeds
- $(1 - p)$ fails, go reverse

Question: **Find final values of all positions**.

- For $p = 0.8$, $\gamma = 1$, $\text{cost} = -4$
- Value of a4 after one iteration?

$$V(s) \leftarrow \left[\arg \max_a \gamma \sum_{s'} P(s'|a)V(s')\right] + R(s)\,.$$

	1	2	3	4
a				76
b	-100			$+100$

Movement probability:

- p succeeds
- $(1-p)$ fails, go reverse

Question: **Find final values of all positions**.

- For $p = 0.8$, $\gamma = 1$, $\text{cost} = -4$
- Value of a4 after one iteration?

$$V(s) \leftarrow \left[\arg \max_a \gamma \sum_{s'} P(s'|a) V(s')\right] + R(s) \,.$$

	1	2	3	4
a				
b	-100			$+100$

Movement probability:

- p succeeds
- $(1 - p)$ fails, go reverse

Question: **Find final values of all positions**.

- For $p = 0.8$, $\gamma = 1$, $\text{cost} = -4$
- Value of a4 after it converges?

$$V(s) \leftarrow \left[\arg \max_a \gamma \sum_{s'} P(s'|a)V(s')\right] + R(s).$$

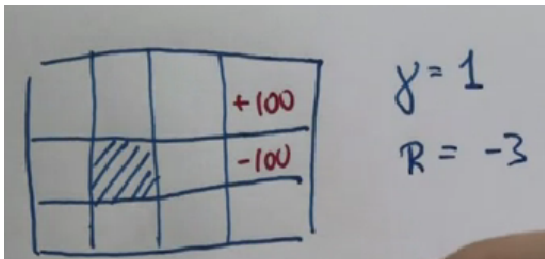|	1	2	3	4
a				95
b	-100			$+100$

Movement probability:
- p succeeds
- $(1 - p)$ fails, go reverse

Question: **Find final values of all positions**.

- For $p = 0.8$, $\gamma = 1$, $\text{cost} = -4$
- Value of a4 after it converges?

$$V(s) \leftarrow \left[\arg\max_a \gamma \sum_{s'} P(s'|a)V(s')\right] + R(s).$$

	1	2	3	4
a				
b	-100			$+100$

Movement probability:

- p succeeds
- $(1 - p)$ fails, go reverse

Question: **Find final values of all positions**.

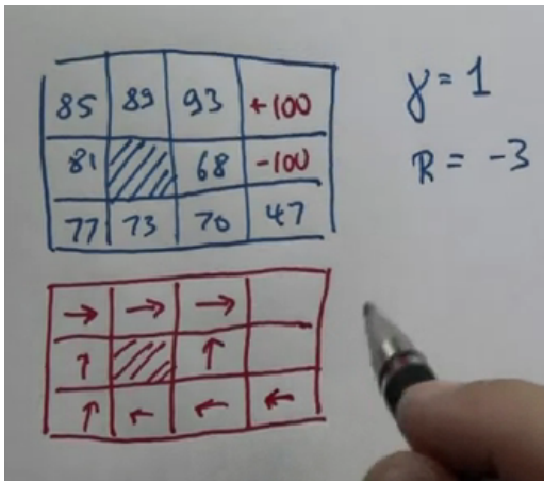- For $p = 0.8$, $\gamma = 1$, $\text{cost} = -4$
- Optimal policy after convergence?

$$V(s) \leftarrow \left[\arg\max_a \gamma \sum_{s'} P(s'|a)V(s') \right] + R(s).$$

	1	2	3	4
a	$\rightarrow$	$\rightarrow$	$\downarrow$	$\downarrow$
b	-100	$\downarrow$	$\rightarrow$	$+100$

Movement probability:
- p succeeds
- $(1-p)$ fails, go reverse

Question: **Find final values of all positions**.

- For $p = 0.8$, $\gamma = 1$, $\text{cost} = -4$
- Optimal policy after convergence?

What happens When Value Iterations Converge?

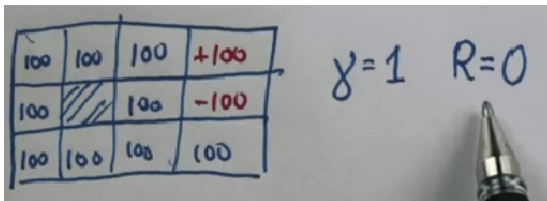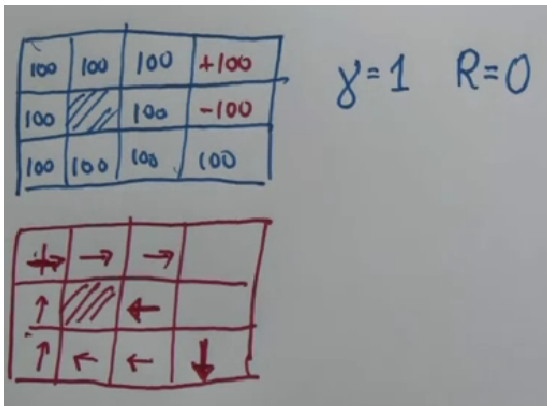$$V(s) \leftarrow \left[\arg \max_a \gamma \sum_{s'} P(s'|a)V(s') \right] + R(s), \quad P(s,a) = 0.8$$

$$V(s) \leftarrow \left[\arg \max_a \gamma \sum_{s'} P(s'|a) V(s')\right] + R(s), \quad P(s, a) = 0.8$$

$$V(s) \leftarrow \left[\arg\max_a \gamma \sum_{s'} P(s'|a)V(s')\right] + R(s), \quad P(s,a) = 0.8$$

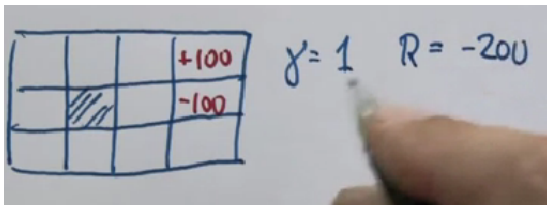$$V(s) \leftarrow \left[\arg\max_a \gamma \sum_{s'} P(s'|a)V(s')\right] + R(s), \quad P(s,a) = 0.8$$

What happens When Value Iterations Converge?

$$V(s) \leftarrow \left[\arg\max_a \gamma \sum_{s'} P(s'|a)V(s')\right] + R(s), \quad P(s,a) = 0.8$$

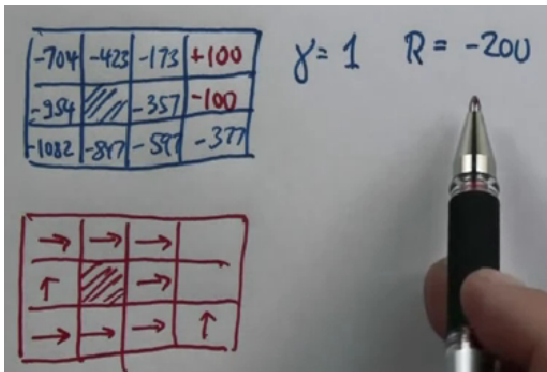$$V(s) \leftarrow \left[\arg\max_a \gamma \sum_{s'} P(s'|a)V(s') \right] + R(s), \quad P(s,a) = 0.8$$

What happens When Value Iterations Converge?

$$V(s) \leftarrow \left[\arg \max_a \gamma \sum_{s'} P(s'|a) V(s') \right] + R(s), \quad P(s,a) = 0.8$$

Exploration and Prediction with MDPs and Reinforcement Learning

Agent Demo

Midterm Evaluations

Please post your evaluations! You have the chance to grade me! :)

Midterm Evaluations

Please post your evaluations! You have the chance to grade me! :)

- Any questions about midterm problems?

Midterm Evaluations

Please post your evaluations! You have the chance to grade me! :)

- Any questions about midterm problems?

- Have a Happy Spring Break!