The Value Iteration Algorithm

Alice Gao

Lecture 15

Based on work by K. Leyton-Brown, K. Larson, and P. van Beek

Outline

Learning Goals

By the end of the lecture, you should be able to

- Describe/trace value iteration for a Markov Decision Process.

CQ: Determine optimal action given true utility

CQ: What is the optimal action in state s_{13}?

(A) Up

(B) Down

(C) Left

(D) Right

	1	2	3	4
1	0.705	0.655	0.611	0.388
2	0.762	X	0.660	-1
3	0.812	0.868	0.918	+1

CQ: Solve system of equations efficiently

CQ: Can we solve this system of equations efficiently?

(A) Yes

(B) No

(C) I don't know

The Bellman equation for $U(s_{11})$:

$$\begin{aligned} U(s_{11}) = -0.04 + \gamma \max[&0.8U(s_{12}) + 0.1U(s_{21}) + 0.1U(s_{11}), \\ &0.9U(s_{11}) + 0.1U(s_{12}), \\ &0.9U(s_{11}) + 0.1U(s_{21}), \\ &0.8U(s_{21}) + 0.1U(s_{12}) + 0.1U(s_{11})]. \end{aligned}$$

CQ: Value iteration

CQ: What is $U_1(s_{23})$?

(A) $(-\infty, 0)$

(B) $[0, 0.25)$

(C) $[0.25, 0.5)$

(D) $[0.5, 0.75)$

(E) $[0.75, 1]$

CQ: Value iteration

CQ: What is $U_2(s_{33})$?

(A) $(-\infty, 0)$

(B) $[0, 0.25)$

(C) $[0.25, 0.5)$

(D) $[0.5, 0.75)$

(E) $[0.75, 1]$

CQ: Value iteration

CQ: What is $U_2(s_{23})$?

(A) $(-\infty, 0)$

(B) $[0, 0.25)$

(C) $[0.25, 0.5)$

(D) $[0.5, 0.75)$

(E) $[0.75, 1]$

CQ: Value iteration

CQ: What is $U_2(s_{32})$?

(A) $(-\infty, 0)$

(B) $[0, 0.25)$

(C) $[0.25, 0.5)$

(D) $[0.5, 0.75)$

(E) $[0.75, 1]$

Revisiting the Learning Goals

By the end of the lecture, you should be able to

- Describe/trace value iteration for a Markov Decision Process.