Week 9: Value Iteration

Monday, April 11, 2022 8:57 AM

Taking the example world:

→ The world contains the set of states $S = \{ s_1, s_2, s_3, sF_1, sF_2 \}$ where s_1 = initial state and, sF_1 and sF_2 are final states:

sF_1	s_1	s_2	s_3	sF_2

→ The world has the following set of Actions $A = \{ \rightarrow, \leftarrow \}$, where:
- $\rightarrow$ = agent moves to left, one cell.
- $\leftarrow$ = agent moves to right, one cell.

→ The reward function $f_R(s, a, s_f) = f_R(s_f)$ only depends on the state that the agent arrives to.

$$f_R(s_f) = \begin{array}{c} sF_1 \\ s_1 \\ s_2 \\ s_3 \\ sF_2 \end{array} \begin{bmatrix} -10 \\ 0 \\ -0.4 \\ -0.4 \\ 10 \end{bmatrix}$$

Now, what we want to solve is the value of $V(S)$, and $V(S)$ is :

$V(sF_1)$ $V(s_1)$ $V(s_2)$ $V(s_3)$ $V(sF_2)$ } all $V(S)$'s ↓ Five $V(S)$'s ↓ each a Bellman's equation

sF_1 — s_1 — s_2 — s_3 — sF_2

Now the excel file, choose sheet WITH $F_\pi(s)$: (Known)

	Transition Model	
	a1=←	a2=→
s1	sF1	s2
s2	s1	s3
s3	s2	sF2
sF1	sF1	s1
sF2	s3	sF2

We also said this is an example of the case where we have a defined politic $f_\pi(s)$

because that row is just input for the V(S) row below (formula)

reward function

		a1=←	a2=→		
	sF1	s1	s2	s3	sF2
fR(s)=	-10	0	-0.4	-0.4	10
fπ(s)=	a1=←	a2=→	a2=→	a2=→	a2=→
V(s)=	0	0	0	0	0
V(s)=	-10	-0.4	-0.4	10	10
γ=	0.9				
	-1.9	-0.076	0.86	1.9	1.9

copy them here

copy the values somewhere else (avoid formula)

This is what we want to find : One V(S) value per state. Start the Value Iteration for each V(S) with a random number, say ∅.

Here, on each cell, we have the Bellman equation for that state.

Since the next state is itself, the formula uses:

$10 = 10 + 0.9 * \emptyset$

Therefore, what we have here, is V(S) of the first iteration.

↓

what we do is put them again where the ∅'s are, since it is our new input

Here, the formula uses $V(S_f)$

$-0.9 = -0.4 + 0.9 * \emptyset$ } Bellman Equation

↳ immediate reward after transition

next state (right) values

} for s_2, it uses s_3 values, and so on.

↳ Do this until the value converges: both V(S) rows are equal. V(S) input row will be the same as the numbers we got from THE ANALYTICAL SOLUTION (solving the linear system).

→ This was a deterministic example. In the case of non-deterministic world, we would need to write $\underbrace{\text{each}}_{\text{each cell of } V(S) \text{ second row.}}$ bellman's equation using the probability distribution.

similar to what we wrote for last section's small example: $S_0 \longrightarrow S_1$

In other words, instead of using in a cell:

$$V(s) = f_R(s, f_\pi(s), s_f) + \gamma V(s_f) \qquad \text{Det}$$

We need to use:

$$V(s) = \sum_{\underbrace{s_f \in S}_{\text{five terms in the cell's sum}}} P_{MT}(s_f \mid s, f_\pi(s)) \left[f_R(s, f_\pi(s), s_f) + \gamma V(s_f) \right] \qquad \text{Non Det}$$

Then, you iterate in the same way.