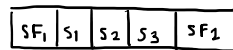


Week 9: Value Iteration

Monday, April 11, 2022 8:57 AM

Taking the example world:

→ The world contains the set of states $S = \{s_1, s_2, s_3, s_{F1}, s_{F2}\}$ where s_1 = initial state and, s_{F1} and s_{F2} are final states:



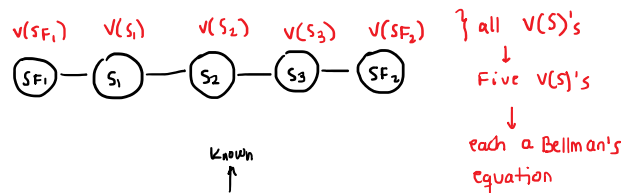
→ The world has the following set of Actions $A = \{\rightarrow, \leftarrow\}$, where:

- \rightarrow = agent moves to left, one cell.
- \leftarrow = agent moves to right, one cell.

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

$$r(s_f) = \begin{matrix} s_{F1} & -10 \\ s_1 & 0 \\ s_2 & -0.4 \\ s_3 & -0.4 \\ s_{F2} & 10 \end{matrix}$$

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



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

Transition Model		
	$a1=\leftarrow$	$a2=\rightarrow$
s_1	s_{F1}	s_2
s_2	s_1	s_3
s_3	s_2	s_{F2}
s_{F1}	s_{F1}	s_1
s_{F2}	s_3	s_{F2}

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

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

copy them here

	$a1=\leftarrow$	$a2=\rightarrow$	
$fr(s)=$	s_{F1}	s_1	s_2
$fr(s)=$	-10	0	-0.4
$fr(s)=$	$a1=\leftarrow$	$a2=\rightarrow$	$a2=\rightarrow$
$V(s)=$	0	0	0
$V(s)=$	-10	-0.4	-0.4
$y=$	0.9	-1.9	-0.076

reward function

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 0.

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

Since the next state is itself, the formula uses:

$$10 = 10 + 0.9 * 0$$

Here, the formula uses $V(s_f)$

$-0.4 = -0.4 + 0.9 * 0$ } Bellman Equation

→ immediate reward after transition

next state (right) values

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

copy the values somewhere else (avoid formula)

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

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

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 each bellman's equation using the probability distribution.
 ↳ each cell of $V(s)$ second row.

similar to what we wrote for last section's small example: $(s_0) \rightarrow (s_1)$
 In other words, instead of using in a cell:

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

We need to use:

$$V(s) = \sum_{\substack{s_f \in S \\ \underbrace{\hspace{1cm}} \\ \text{five terms in} \\ \text{the cell's sum}}} p_{M\pi}(s_f | s, f_\pi(s)) [f_R(s, f_\pi(s), s_f) + \gamma V(s_f)] \quad \text{Non Det}$$

Then, you iterate in the same way.