Week 6: Homework

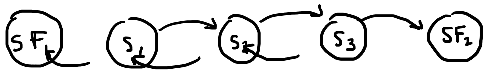
Saturday, March 12, 2022 10:40 AM

⑤ 1 dimensional World

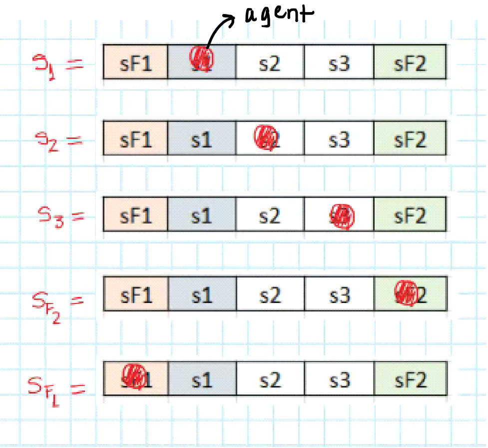
$S = \{ s_1, s_2, s_3, sF_1, sF_2$

initial state s_1 : we always begin here.

final states : whenever you reach a final state, the program ends. You cannot go out of it.



when we say " state s_i" we mean the agent is in that position in the world, and so on :



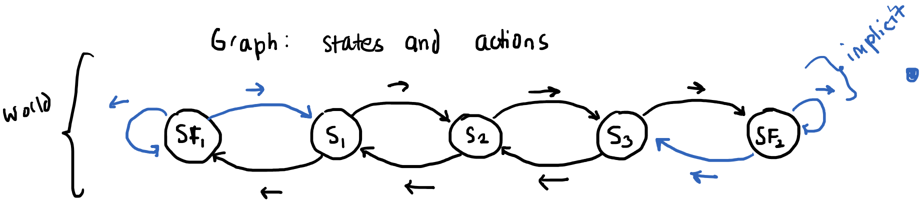
The same happens with the reward diagram

↳ does not depend on which state you came from, only on the s_f (final state after an action)

i.e. no matter from where I arrive to $\underbrace{s_2}_{sf}$, the reward is always −0.4

a) Build the graph of the world

Graph: states and actions



● Not possible since final state, but World must include all actions even though it's not possible

Transition model: deterministic, since given an action a, the agent always moves 1 unit to the same state.

b) Write the transition function $f_{MT}(s,a)$

$$f_{MT}(s,a) = \begin{array}{c|cc} & \leftarrow & \rightarrow \\ s_1 & SF_1 & S_2 \\ s_2 & S_1 & S_3 \\ s_3 & S_2 & SF_2 \\ SF_1 & SF_1 & S_1 \\ SF_2 & S_3 & SF_2 \end{array}$$

If you didn't put transition out of SF_1, this should be not defined and it's ok.

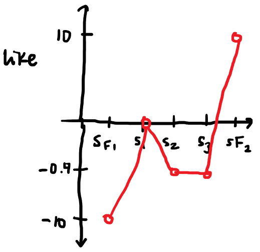
c) Write the reward function (function of one variable)

$$f_R(s_f) = \begin{array}{c|c} s_1 & 0 \\ s_2 & -0.4 \\ s_3 & -0.9 \\ SF_1 & -10 \\ SF_2 & 10 \end{array}$$

s_f

this function represents the reward given at the end of each transition

or also like



d) 2 functions of action $f_\pi(s)$: make two politics

I. from s_1 to final state SF_2

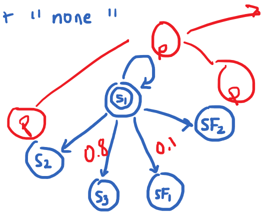
$$f_\pi(s) = \begin{array}{c|c} s_1 & \rightarrow \\ s_2 & \rightarrow \\ s_3 & \rightarrow \\ S_{F_1} & \leftarrow \\ SF_2 & \rightarrow \end{array}$$

so that it stays in the final state it's ok to put " none "

$$= \begin{array}{|c|c|c|c|c|} SF_2 & s_1 & s_2 & s_3 & SF_2 \\ \hline \leftarrow & \rightarrow & \rightarrow & \rightarrow & \rightarrow \end{array}$$

Collect

$0.3a_3$



we know only that the sum is 0.1

since

$1-(0.8+0.1)$

$= 0.1$

II. from s_1 to final state SF_1

$$f_\pi(s) = \begin{array}{c|c} s_1 & \leftarrow \\ s_2 & \leftarrow \\ s_3 & \leftarrow \\ S_{F_1} & \leftarrow \\ S_{F_2} & \rightarrow \end{array}$$

since these never happens, no matter which actions you put here.

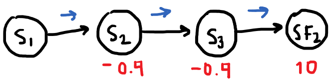
Let's calculate the accumulated reward function:

$\tau = s_1, s_2, s_3, SF_2$

or

The trajectory

$\tau =$



(gamma) $\gamma = 0.8$, let's say.

Thus, for τ,

$$\begin{aligned} f_{AR}(\tau) &= -0.4 + \gamma(-0.4) + \gamma^2(10) \\ &= -0.4 + (0.8)(-0.4) + (0.8)^2(10) \\ &= -0.4 + (-0.32) + 6.4 \\ &= \boxed{5.68} \end{aligned}$$

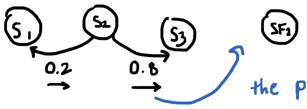
⑥

→ = to the right with $p=0.8$

to the left with $p=0.2$

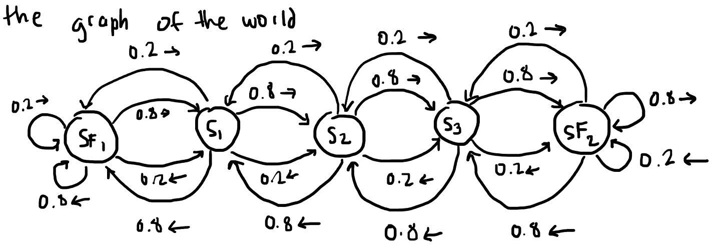
8 out of 10 times, → moves to the right

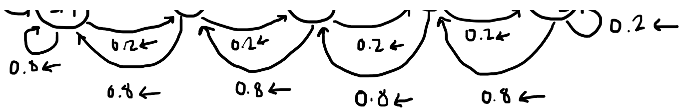
=



the probability of arriving to SF_2 from s_1 by taking → is ∅

a) Build the graph of the world





b) Transition function $P_{MT}(s_f \mid s,a)$ ← 3D matrix → stochastic World

$$P_{MT}(s_f \mid s,a) = \begin{array}{c|cc|} & \leftarrow & \rightarrow \\ s_1 & 0 & 0 \\ s_2 & 0.8 & 0.2 \\ s_3 & 0 & 0 \\ sF_1 & 0.2 & 0.8 \\ sF_2 & 0 & 0 \end{array} \quad \begin{array}{c|cc|} & \leftarrow & \rightarrow \\ s_1 & 0.2 & 0.8 \\ s_2 & 0 & 0 \\ s_3 & 0.8 & 0.2 \\ sF_1 & 0 & 0 \\ sF_2 & 0 & 0 \end{array} \quad \begin{array}{c|cc|} & \leftarrow & \rightarrow \\ s_1 & 0 & 0 \\ s_2 & 0.2 & 0.8 \\ s_3 & 0 & 0 \\ sF_1 & 0 & 0 \\ sF_2 & 0.8 & 0.2 \end{array} \quad \begin{array}{c|cc|} & \leftarrow & \rightarrow \\ s_1 & 0.8 & 0.2 \\ s_2 & 0 & 0 \\ s_3 & 0 & 0 \\ sF_1 & 0.8 & 0.2 \\ sF_2 & 0 & 0 \end{array} \quad \begin{array}{c|cc|} & \leftarrow & \rightarrow \\ s_1 & 0 & 0 \\ s_2 & 0 & 0 \\ s_3 & 0.2 & 0.8 \\ sF_1 & 0 & 0 \\ sF_2 & 0.2 & 0.8 \end{array}$$

$s_f = s_1$ $\quad s_f = s_2$ $\quad s_f = s_3$ $\quad s_f = sF_1$ $\quad s_f = sF_2$

c) Write the reward function

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

s_f

(same as 5.c)

d) 2 action functions $f_\pi(s)$

- look for a politic that most of the time arrives to sF_2
  ↳ take the case of 0.8 prob (→)

I. $$f_\pi(s) = \begin{array}{c|c|} s_1 & \rightarrow \\ s_2 & \rightarrow \\ s_3 & \rightarrow \\ sF_1 & \leftarrow \\ sF_2 & \rightarrow \end{array}$$

← so that it stays in SF_1

II. $$f_\pi(s) = \begin{array}{c|c|} s_1 & \leftarrow \\ s_2 & \leftarrow \\ s_3 & \leftarrow \\ sF_1 & \leftarrow \\ sF_2 & \rightarrow \end{array}$$

→ in case of ← going to s_2, we go back and so on

e) Probability of arriving to sF_2 and sF_1 using the last 4. politics.

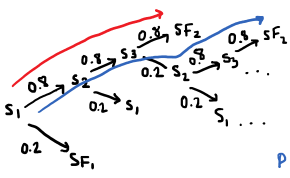
all probabilities are independent
↳ all variables are independent
(decision trees)

I.

$$P = \underbrace{(0.8)(0.8)(0.8)}_{\text{since they are independent from history (indep. events) and so we multiply to get their 'together' distribution}} = \boxed{0.512}$$

→ probability of this path only

→ half of times this happens
↓
the other half is all the other options in the tree



$$P = (0.8)(0.8)(0.2)(0.8)(0.8) = 0.08$$

If we do not find a pattern to the convergence, just .512