Homework 02: Review

Saturday, March 26, 2022 10:41 AM

(2) since $f_R(s_t)$ -> the reward has nothing to do with the action done or the previous state.

$$\tau_1 = s_1 \xrightarrow{1} s_2 \xrightarrow{-1} s_3 \xrightarrow{2} s_1 \xrightarrow{1} s_2 \xrightarrow{2} s_1$$

even if there were two actions that $s_1 \to s_2$, the reward is the same

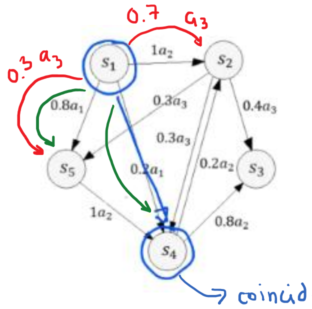
Thus, one path

$$f_{RA}(\tau_1) = 1 + \gamma(-1) + \gamma^2(2) + \gamma^3(1) + \gamma^4(2)$$

if they asked

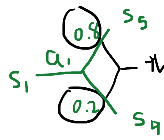
$f_{RA}(s_1) = V(s_1)$ = average acc reward of all trayectories starting on s_1

4.



→ coincidentally, on every state there is one action that takes you to another state.

The probabilities of transition are not used since we are calculating the f_{AR} of just one trayectory.

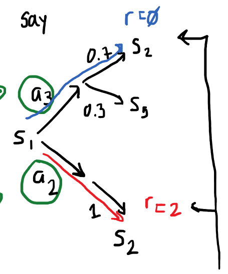


they sum 1

Let's suppose what's on red

now we have two actions that take us from s_1 to s_2 (a_3 and a_2)

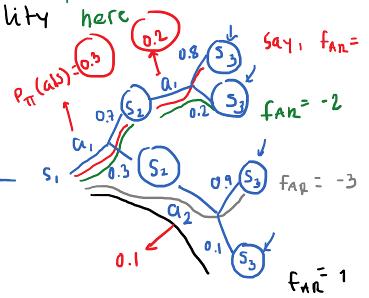
Thus the decision of which action to take can be decided based on a distribution. Let's say



$\tau_1 = s_1, s_2$ → we have two possible rewards

Thus the politic must be stochastic $\underbrace{P_\pi(a|s)}$ this defines the probability here

tree of trajectories ←



say, $f_{AR} = 3$ → this will be multiplied by the probabilities (0.3)(0.2) so that it can be summed with the others, instead of doing the average like before.

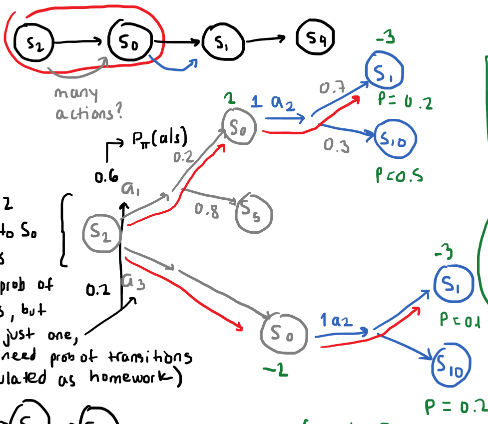
$$\overline{f_{AR}} = \underbrace{(0.3)(0.7)(0.2)(0.2)}_{\text{prob of happening}}\overset{f_{AR}}{(3)} + \ldots$$

Let's say

$\tau_1 = s_2, s_0, s_1, s_4$ with the politic $P_\pi(a|s)$ ← based on state

Let's say

$T_1 = s_2, s_0, s_1, s_4$ with the politic $P_\pi(a|s)$ ← based on state



thus,

We have 2 actions to s_0 thus f_{AR} is using prob of transitions, but if we had just one, you don't need prob of transitions (calculated as homework)

$f_{AR} = (2 + \gamma(-3))(P/(0.2+0.1))$

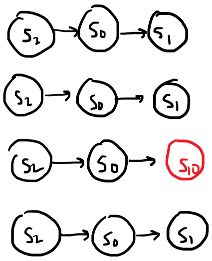
where $P = (0.6)(0.2)(1)(0.7)$

$f_{AR} = (-2 + \gamma(-3))(P/\underbrace{(0.2+0.1)})$ → sum of the ones that arrive at the state I want

then sum to have the f_{AR} in this trayectory

which is $\neq V(s_2)$

from the whole tree { all possible paths sum 1 (0.2+0.5+0.1+0.2)



Class:

Let's use M_T non deterministic Bellman's Equation

M_T NO determinista

$V(s) = \overline{f_{RA}(\tau)}\Big|_{s(0)=s}$ this can be written as this

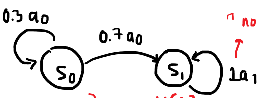
$= \overline{f_R(s,a,s_f) + \gamma V(s_f)}$ also like this

$= \sum_{s_f \in S} P_{M_T}(s_f|s,a)[f_R(s,a,s_f) + \gamma V(s_f)]$

con: $a = f_\pi(s)$ plug this

which result in

$V(s) = \sum_{s_f \in S} P_{M_T}\left(s_f|s, f_\pi(s)\right)[f_R(s, f_\pi(s), s_f) + \gamma V(s_f)]$ } Bellman's equation in a general way.



$V(s_0) = -2.65$ all these infinite trajectories starting on s_0 and averaged are -2.65

$V(s_1) = \emptyset$

we used to say "there are ∞ trayectories, we can't calculate" now we can

→ $s_f \sim P_{MT}(s_f|s,a) =$

$s_f = s_0$:

	a_0	a_1
s_0	0.3	0
s_1	0	0

$s_f = s_1$:

	a_0	a_1
s_0	0.7	0
s_1		1

→ Reward func: $f_R(s,a,s_f) =$

$s_f = s_0$:

	a_0	a_1
s_0	0	1
s_1	1	2

$s_f = s_1$:

	a_0	a_1
s_0	-3	-1
s_1	1	0

} reward defined arbitrarily (it should teach something)

→ Politic func: $P_\pi(s_f) =$

s_0	a_0
s_1	a_1

with $\gamma = 0.7$

Thus, there must be two $V(s) = V(s_0)$ and $V(s_1)$ (one for each state)

→ 0, 1 all final states (all states are final states)

$$V(s_0) = \sum_{i=0}^{1} P_{MT}(s_i|s_0,a_0)[f_R(s_0,a_0,s_i) + \gamma V(s_i)]$$

$$= \underbrace{P_{MT}(s_0|s_0,a_0)}_{0.3 \text{ from } P_{MT}}[\underbrace{f_R(s_0,a_0,s_0)}_{\emptyset \text{ from } f_R} + \gamma V(s_0)] + \underbrace{P_{MT}(s_1|s_0,a_0)}_{0.7 \text{ from } P_{MT}}[\underbrace{f_R(s_0,a_0,s_1)}_{-3} + \gamma V(s_1)]$$

Plugging known values:

$$V(s) = 0.3[\emptyset + 0.7V(s_0)] + 0.7[(-3) + \gamma V(s_1)]$$

$$= 0.21\, V(S_0) - 2.1 + 0.49\, V(S_1)$$

Thus,

$$V(S_0) - 0.21\, V(S_0) = -2.1 + 0.49\, V(S_1)$$

Giving out the first equation of a 2x2 system

$$\boxed{0.79\, V(S_0) = -2.1 + 0.49\,(S_1)} \quad (1)$$

Let's calculate eq 2

$$V(S_1) = \sum_{i=0}^{1} P_{MT}(S_i \mid S_1, a_1)\left[f_R(S_1, a_1, S_i) + \gamma V(S_i)\right]$$

$$= \underbrace{P_{MT}(S_0 \mid S_1, a_1)}_{\emptyset}\left[\underbrace{f_R(S_1, a_1, S_0)}_{\emptyset} + \gamma\, v(S_0)\right] + \underbrace{P_{MT}(S_1 \mid S_1, a_1)}_{1}\left[\underbrace{f_R(S_1, a_1, S_1)}_{\emptyset} + \gamma\, v(S_1)\right]$$

$$v(S_1) = 1\left[0.7\, v(S_1)\right]$$

$$v(S_1) - 0.7\, v(S_1) = 0$$

$$0.3\, V(S_1) = 0$$

$$v(S_1) = \frac{0}{0.3}$$

$$\boxed{v(S_1) = \emptyset \quad (2)}$$ 2nd bellman's eq

Thus, the system is:

$$0.79\, v(S_0) = -2.1 + \cancel{0.49(\emptyset)}$$

$$v(S_0) = -\frac{2.1}{0.79} = -2.65$$

Which we now go to the graph and complete.