Homework 02: Accumulated Reward

Mariana Ávalos Arce
0197495

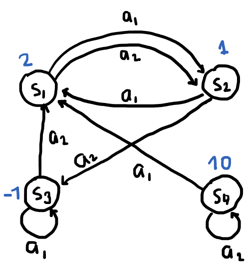
Tuesday, March 22, 2022 10:54 AM

① Given the world defined by the following transition function $f_{MT}(s,a)$, the reward function $f_R(s, a, s_f) = f_R(s_f)$ and $\gamma = 0.9$:

$$f_{M_T}(s,a) = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \\ s_4 \end{array}\begin{array}{c} \begin{array}{cc} a_1 & a_2 \end{array} \\ \begin{bmatrix} s_2 & s_2 \\ s_1 & s_3 \\ s_3 & s_1 \\ s_1 & s_4 \end{bmatrix}\end{array} \qquad f_R(s_f) = \begin{array}{c} s_1 \\ s_2 \\ s_3 \\ s_4 \end{array}\begin{bmatrix} 2 \\ 1 \\ -1 \\ 10 \end{bmatrix}$$

ⓐ Calculate the accumulated reward function $f_{AR}(\tau_1)$ for the trajectory :
$\tau_1 = s_1, s_2, s_3, s_1, s_2, s_1$

Solution



$A = \{a_1\ a_2\}$ $S = \{s_1\ s_2\ s_3\ s_4\}$

$\tau_1 \rightarrow$ $s_1 \xrightarrow[a_1,a_2]{} \overset{1}{s_2} \xrightarrow[a_2]{} \overset{-1}{s_3} \xrightarrow[a_2]{} \overset{2}{s_1} \xrightarrow[a_1,a_2]{} \overset{1}{s_2} \xrightarrow[a_1]{} \overset{2}{s_1}$

$\gamma = 0.9$

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

$= 1 + (0.9)(-1) + (0.9)^2(2) + (0.9)^3(1) + (0.9)^4(2)$

$= 3.76$

$\boxed{f_{AR}(\tau_1) = 3.76}$

ⓑ Calculate the accumulated reward function $f_{AR}(\tau_1)$ for the trajectory :
$\tau_2 = s_3, s_1, s_2, s_3$

Solution

$s_3 \rightarrow \overset{2}{s_1} \rightarrow \overset{1}{s_2} \rightarrow \overset{-1}{s_3}$

with $\gamma = 0.9$

$f_{AR}(\tau_2) = 2 + \gamma(1) + \gamma^2(-1)$

$= 2 + 0.9(1) + (0.9)^2(-1)$

$\boxed{f_{AR}(\tau_2) = 2.09}$

ⓒ Calculate the accumulated reward function $f_{AR}(\tau_1)$ for the trajectory :
$\tau_3 = s_2, s_1, s_2$

Solution

$s_2 \rightarrow \overset{2}{s_1} \rightarrow \overset{1}{s_2}$

$f_{AR}(\tau_3) = 2 + \gamma(1)$

$= 2 + 0.9(1)$

$= 2 + 0.9$

$\boxed{f_{AR}(\tau_3) = 2.9}$

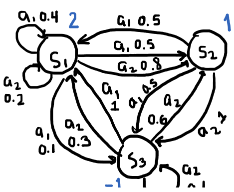
② Given the world defined by the transition function $P_{MT}(s_f \mid s, a)$, the reward function $f_R(s_f, s, a) = f_R(s_f)$ and $\gamma = 0.6$:

$$f_{M_T}(s,a) = \begin{array}{c} \\ \\ s_1 \\ s_2 \\ s_3 \end{array}\begin{array}{c} s_f = s_1 \\ \begin{array}{cc} a_1 & a_2 \end{array} \\ \begin{bmatrix} 0.4 & 0.2 \\ 0.5 & 0 \\ 1 & 0.3 \end{bmatrix}\end{array} \quad \begin{array}{c} \\ \\ s_1 \\ s_2 \\ s_3 \end{array}\begin{array}{c} s_f = s_2 \\ \begin{array}{cc} a_1 & a_2 \end{array} \\ \begin{bmatrix} 0.5 & 0.8 \\ 0 & 0 \\ 0 & 0.6 \end{bmatrix}\end{array} \quad \begin{array}{c} \\ \\ s_1 \\ s_2 \\ s_3 \end{array}\begin{array}{c} s_f = s_3 \\ \begin{array}{cc} a_1 & a_2 \end{array} \\ \begin{bmatrix} 0.1 & 0 \\ 0.5 & 1 \\ 0 & 0.1 \end{bmatrix}\end{array} \qquad f_R(s_f) = \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array}\begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$$

ⓐ Calculate the accumulated reward function $f_{AR}(\tau_1)$ for the trajectory :
$\tau_1 = s_1, s_2, s_3, s_1, s_2, s_1$

Solution

$S = \{s_1\ s_2\ s_3\}$ $A = \{a_1\ a_2\}$



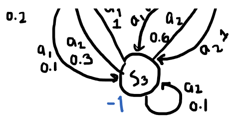
$\tau_1 =$ $s_1 \rightarrow \overset{1}{s_2} \rightarrow \overset{-1}{s_3} \rightarrow \overset{2}{s_1} \rightarrow \overset{1}{s_2} \rightarrow \overset{2}{s_1}$

with $\gamma = 0.6$

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

$= 1 + (0.6)(-1) + (0.6)^2(2) + (0.6)^3(1) + (0.6)^4(2)$

$\boxed{f_{AR}(\tau_1) = 1.59}$



$$f_{AR}(\tau_1) = 1 + \gamma(-1) + \gamma^2(2) + \gamma^3(1) + \gamma^4(2)$$
$$= 1 + (0.6)(-1) + (0.6)^2(2) + (0.6)^3(1) + (0.6)^4(2)$$
$$\boxed{f_{AR}(\tau_1) = 1.59}$$

(b) Calculate the accumulated reward function $f_{AR}(\tau_2)$ for the trajectory :
$\tau_2 = S_3, S_1, S_2, S_3$

Solution

$$S_3 \xrightarrow{2} S_1 \xrightarrow{1} S_2 \xrightarrow{-1} S_3$$

with $\gamma = 0.6$

$$f_{AR}(\tau_2) = 2 + \gamma(1) + \gamma^2(-1)$$
$$= 2 + (0.6)(1) + (0.6)^2(-1)$$
$$\boxed{f_{AR}(\tau_2) = 2.24}$$

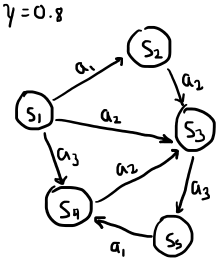
(c) Calculate the accumulated reward function $f_{AR}(\tau_3)$ for the trajectory :
$\tau_3 = S_3, S_1, S_2$

Solution

$$S_3 \xrightarrow{2} S_1 \xrightarrow{1} S_2$$

$$f_{AR}(\tau_3) = 2 + \gamma(1)$$
$$= 2 + 0.6(1)$$
$$\boxed{f_{AR}(\tau_3) = 2.6}$$

(3) Given the world defined by the following graph, the reward function $f_R(s_f, s, a)$ and $\gamma = 0.8$



$$f_R(s,a,s_f) = \underset{s_f = s_1}{\begin{array}{c|ccc} & a_1 & a_2 & a_3 \\ \hline s_1 & 0 & 0 & 0 \\ s_2 & 0 & 0 & 0 \\ s_3 & 0 & 0 & 0 \\ s_4 & 0 & 0 & 0 \\ s_5 & 0 & 0 & 0 \end{array}} \quad \underset{s_f = s_2}{\begin{array}{c|ccc} & a_1 & a_2 & a_3 \\ \hline s_1 & -2 & 0 & 0 \\ s_2 & 0 & 0 & 0 \\ s_3 & 0 & 0 & 0 \\ s_4 & 0 & 0 & 0 \\ s_5 & 0 & 0 & 0 \end{array}} \quad \underset{s_f = s_3}{\begin{array}{c|ccc} & a_1 & a_2 & a_3 \\ \hline s_1 & 0 & 5 & 0 \\ s_2 & 0 & 4 & 0 \\ s_3 & 0 & 0 & 0 \\ s_4 & 0 & -1 & 0 \\ s_5 & 0 & 0 & 0 \end{array}} \quad \underset{s_f = s_4}{\begin{array}{c|ccc} & a_1 & a_2 & a_3 \\ \hline s_1 & 0 & 0 & -3 \\ s_2 & 0 & 0 & 0 \\ s_3 & 0 & 0 & 0 \\ s_4 & 0 & 0 & 0 \\ s_5 & 1 & 0 & 0 \end{array}} \quad \underset{s_f = s_5}{\begin{array}{c|ccc} & a_1 & a_2 & a_3 \\ \hline s_1 & 0 & 0 & 0 \\ s_2 & 0 & 0 & 0 \\ s_3 & 0 & 0 & -6 \\ s_4 & 0 & 0 & 0 \\ s_5 & 0 & 0 & 0 \end{array}}$$

(a) Calculate the accumulated reward function $f_{AR}(\tau_1)$ for the trajectory :
$\tau_1 = S_1, S_2, S_3, S_5, S_4, S_3, S_5$

Solution

$$S_1 \xrightarrow{-2} S_2 \xrightarrow{4} S_3 \xrightarrow{-6} S_5 \xrightarrow{1} S_4 \xrightarrow{-1} S_3 \xrightarrow{-6} S_5$$

with $\gamma = 0.8$

$$f_{AR}(\tau_1) = -2 + \gamma(4) + \gamma^2(-6) + \gamma^3(1) + \gamma^4(-1) + \gamma^5(-6)$$
$$= -2 + (0.8)(4) + (0.8)^2(-6) + (0.8)^3(1) + (0.8)^4(-1) + (0.8)^5(6)$$
$$\boxed{f_{AR}(\tau_1) = -0.57}$$

(b) Calculate the accumulated reward function $f_{AR}(\tau_2)$ for the trajectory :
$\tau_2 = S_1, S_3, S_5, S_4$

Solution

$$S_1 \xrightarrow{5} S_3 \xrightarrow{-6} S_5 \xrightarrow{1} S_4$$

with $\gamma = 0.8$

$$f_{AR}(\tau_2) = 5 + \gamma(-6) + \gamma^2(1)$$
$$= 5 + (0.8)(-6) + (0.8)^2(1)$$
$$\boxed{f_{AR}(\tau_2) = 0.84}$$

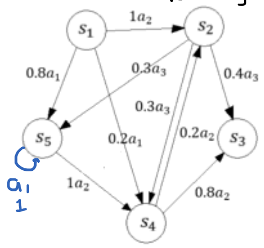
(c) Calculate the accumulated reward function $f_{AR}(\tau_3)$ for the trajectory :
$\tau_3 = S_4, S_3, S_5$

$$S_4 \xrightarrow{-1} S_3 \xrightarrow{-6} S_5$$

with $\gamma = 0.8$

$$f_{AR}(\tau_3) = -1 + \gamma(6)$$
$$= -1 + (0.8)(6)$$
$$= -1 + 4.8$$
$$\boxed{f_{AR}(\tau_3) = 3.8}$$

(4) Given the world defined by the following graph where $0.8a_1$ means a_1 with probability 0.8, and so on, and with the following reward function and $\gamma = 0.7$



$$f_R(s,a,s_f) = \begin{array}{c}s_f = s_1\\ \begin{array}{c|ccc} & a_1 & a_2 & a_3 \\ \hline s_1 & 9 & 0 & 0 \\ s_2 & 0 & 0 & 0 \\ s_3 & 0 & 1 & 0 \\ s_4 & 0 & 0 & -2 \\ s_5 & 2 & 0 & 0 \end{array}\end{array} \quad \begin{array}{c}s_f = s_2\\ \begin{array}{c|ccc} & a_1 & a_2 & a_3 \\ \hline s_1 & -2 & 2 & 0 \\ s_2 & 0 & 0 & 0 \\ s_3 & 0 & 0 & 0 \\ s_4 & 0 & -3 & 0 \\ s_5 & 0 & 0 & 0 \end{array}\end{array} \quad \begin{array}{c}s_f = s_3\\ \begin{array}{c|ccc} & a_1 & a_2 & a_3 \\ \hline s_1 & 0 & 5 & 0 \\ s_2 & 0 & 4 & -1 \\ s_3 & 0 & 0 & 0 \\ s_4 & 0 & -1 & 0 \\ s_5 & 0 & 0 & 0 \end{array}\end{array} \quad \begin{array}{c}s_f = s_4\\ \begin{array}{c|ccc} & a_1 & a_2 & a_3 \\ \hline s_1 & 1 & 0 & -3 \\ s_2 & 0 & 0 & 2 \\ s_3 & 0 & 0 & 0 \\ s_4 & 0 & 0 & 0 \\ s_5 & 1 & -1 & 0 \end{array}\end{array} \quad \begin{array}{c}s_f = s_5\\ \begin{array}{c|ccc} & a_1 & a_2 & a_3 \\ \hline s_1 & 3 & 0 & 0 \\ s_2 & 0 & 0 & 0 \\ s_3 & 0 & 0 & -6 \\ s_4 & 0 & 0 & 0 \\ s_5 & 0 & 0 & 0 \end{array}\end{array}$$

(a) Calculate the accumulated reward function $f_{AR}(\tau_1)$ for the trajectory:

$\tau_1 = s_1, s_4, s_2, s_5, s_4, s_3$ no transition

Solution

$s_1 \xrightarrow[a_1]{0.2 \quad 1} s_4 \xrightarrow[a_2]{0.2 \quad -3} s_2 \xrightarrow[a_3]{0.3 \quad 0} s_5 \xrightarrow[a_1]{0 \quad 1} s_4 \xrightarrow[a_2]{0.8 \quad -1} s_3$ No transition, not a path

$s_1 \xrightarrow[a_3]{0 \quad -3} s_4 \xrightarrow[a_2]{0.2 \quad -3} s_2 \xrightarrow[a_3]{0.3 \quad 0} s_5 \xrightarrow[a_2]{1 \quad -1} s_4 \xrightarrow[a_2]{0.8 \quad -1} s_3$ no transition from $s_1 \to s_4$ with a_3

$s_1 \xrightarrow[a_1]{0.2 \quad 1} s_4 \xrightarrow[a_2]{0.2 \quad -3} s_2 \xrightarrow[a_3]{0.3 \quad 0} s_5 \xrightarrow[a_2]{1 \quad -1} s_4 \xrightarrow[a_2]{0.8 \quad -1} s_3$ $f_{AR} = -1.68$

$s_1 \xrightarrow[a_3]{0 \quad -3} s_4 \xrightarrow[a_2]{0.2 \quad -3} s_2 \xrightarrow[a_3]{0.3 \quad 0} s_5 \xrightarrow[a_1]{0 \quad 1} s_4 \xrightarrow[a_2]{0. \quad -1} s_3$ no transition, not a path

$$f_{AR} = 1 + (0.7)(-3) + (0.7)^2(0) + 0.7^3(-1) + 0.7^4(-1) = \boxed{-1.68}$$

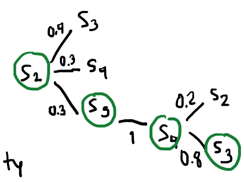
(b) Calculate the accumulated reward function $f_{AR}(\tau_2)$ for the trajectory:

$\tau_2 = s_2, s_5, s_4, s_3$

Solution

$s_2 \xrightarrow[a_3]{0.3 \quad 0} s_5 \xrightarrow[a_1]{0 \quad 1} s_4 \xrightarrow[a_2]{0.9 \quad -1} s_3$ this transition from s_5 to s_4 doesn't exist: this path is not possible

$s_2 \xrightarrow[a_3]{0.3 \quad 0} s_5 \xrightarrow[a_2]{1 \quad -1} s_4 \xrightarrow[a_2]{0.8 \quad -1} s_3$



$$f_{AR} = 0 + (0.7)(-1) + (0.7)^2(-1)$$
$\boxed{f_{AR} = -1.19}$ with 0.24 of probability

(c) Calculate the accumulated reward function $f_{AR}(\tau_2)$ for the trajectory:

$\tau_3 = s_4, s_2, s_3$

Solution

$s_4 \xrightarrow[a_2]{0.2 \quad -3} s_2 \xrightarrow[a_2]{0 \quad 4} s_3$ no transition, not possible

$s_4 \xrightarrow[a_2]{0.2 \quad -3} s_2 \xrightarrow[a_3]{0.4 \quad -1} s_3$ $f_{AR} = -3 + (0.7)(-1) = \boxed{-3.7}$

with 0.08 probability

(5) El mundo tiene el siguiente conjunto de estados S={s1, s2, s3, sF1, sF2} donde s1=estado inicial y, sF1 y sF2 son estados terminales:

sF1	s1	s2	s3	sF2

El mundo tiene el siguiente conjunto de acciones A={→, ←} donde:

→=Agente se mueve a la derecha una sola celda

←=Agente se mueve a la izquierda una sola celda

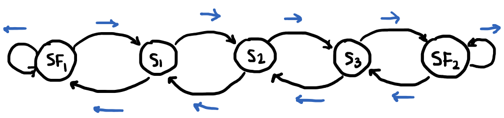
La función de recompensa $f_R(s, a, s_f) = f_R(s_f)$ solo depende del estado al que el Agente llega y esta definida como:

-10	0	-0.4	-0.4	10

Es decir, si el agente transiciona de s1 a s2 entonces recibe la recompensa -0.4 que esta definide en el estado s2. El agente tiene la siguiente función de acción $f_\pi(s)$:

$$f_\pi(s) = \begin{array}{c} s_1 \\ s_2 \\ s_3 \\ s_{F1} \\ s_{F2} \end{array} \begin{bmatrix} \rightarrow \\ \rightarrow \\ \rightarrow \\ \leftarrow \\ \rightarrow \end{bmatrix}$$

(a) Build the graph of the world



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

$$f_{MT}(s,a) = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \\ sF_1 \\ sF_2 \end{array} \begin{array}{c} \leftarrow \quad \rightarrow \\ \begin{bmatrix} sF_1 & s_2 \\ s_1 & s_3 \\ s_2 & sF_2 \\ sF_1 & s_1 \\ s_3 & sF_2 \end{bmatrix} \end{array}$$

(c) Build all the possible trajectories from the initial state s_1 given the action function $f_\pi(s)$ that go to a final state, either sF_1 or sF_2

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

$\tau_1 = S_1, S_2, S_3, SF_2$

$\tau_2 = S_1, S_2, S_3, SF_2, S_3, SF_2$

(d) Build all the possible trajectories from the state s_2, given the action function $f_\pi(s)$ that go to a final state, either sF_1 or sF_2

$\tau_1 = S_2, S_3, SF_2$

$\tau_2 = S_2, S_3, SF_2, S_3, SF_2$

(e) Build all the possible trajectories from the state s_3, given the action function $f_\pi(s)$ that go to a final state, either sF_1 or sF_2

$\tau_1 = S_3, SF_2$

$\tau_2 = S_3, SF_2, S_3, SF_2$

(f) Calculate the accumulated reward of every possible trajectory in c, d, e using $\gamma = 0.7$.

SF_1	S_1	S_2	S_3	SF_2
-10	0	-0.4	-0.4	10

$s(0) = S_1$

$\tau_1 = S_1, S_2, S_3, SF_2$ $\quad f_{RA} = -0.4 + (0.7)(-0.4) + (0.7)^2(10) = 4.22$

$\tau_2 = S_1, S_2, S_3, SF_2, S_3, SF_2$ $\quad f_{RA} = -0.4 + (0.7)(-0.4) + (0.7)^2(10) + (0.7)^3(-0.4) + (0.7)^4(10) = 6.48$

$s(0) = S_2$

$\tau_1 = S_2, S_3, SF_2$ $\quad f_{RA} = -0.4 + (0.7)(10) = -0.4 + 7 = 6.6$

$\tau_2 = s_2, s_3, sF_2, s_3, sF_2 \qquad f_{RA} = -0.4 + (0.7)(10) + (0.7)^2(0.4) + (0.7)^3(10) = 9.83$

$s(0) = s_3$

$\tau_1 = s_3, sF_2 \qquad f_{RA} = 10$

$\tau_2 = s_3, sF_2, s_3, sF_2 \qquad f_{RA} = 10 + (0.7)(-0.4) + (0.7)^2(10) = 14.62$

6. El mundo tiene el siguiente conjunto de estados S={s1, s2, s3, sF1, sF2} donde s1=estado inicial y, sF1 y sF2 son estados terminales:

sF1	s1	s2	s3	sF2

El mundo tiene el siguiente conjunto de acciones A={→, ←} donde:

→=Agente se mueve a la derecha una sola celda con probabilidad 0.8 y se mueve una sola celda a la izquierda con probabilidad 0.2

←=Agente se mueve a la izquierda una sola celda con probabilidad 0.8 y se mueve una sola celda a la derecha con probabilidad 0.2

La función de recompensa $f_R(s,a,s_f) = f_R(s_f)$ solo depende del estado al que el Agente llega y esta definida como:

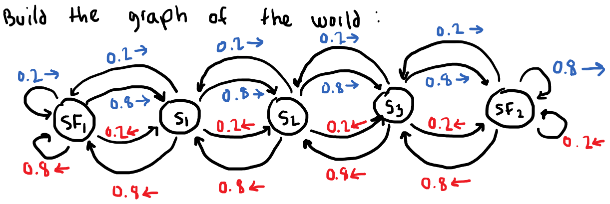
-10	0	-0.4	-0.4	10

Es decir, si el agente transiciona de s1 a s2 entonces recibe la recompensa -0.4 que esta definide en el estado s2.

El agente tiene la siguiente función de acción $f_\pi(s)$:

$$f_\pi(s) = \begin{array}{c} s_1 \\ s_2 \\ s_3 \\ s_{F1} \\ s_{F2} \end{array} \begin{bmatrix} \rightarrow \\ \rightarrow \\ \rightarrow \\ \leftarrow \\ \rightarrow \end{bmatrix}$$

a) Build the graph of the world:

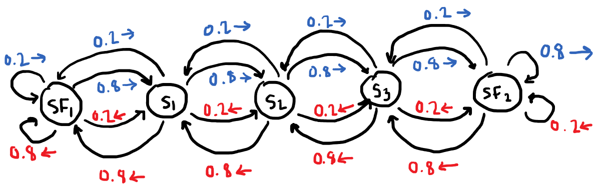


b) Write the transition function $P_{MT}(s_f | s,a)$

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

c. Construya todas las trayectorias posibles a partir del estado inicial s1 dada la función de acción $f_\pi(s)$ que lleven a un estado final ya sea sF1 o sF2
(NOTA: Dado que es un número infinito de trayectorias solo escriba 10)

$$f_\pi(s) = \begin{array}{c} s_1 \\ s_2 \\ s_3 \\ s_{F1} \\ s_{F2} \end{array} \begin{bmatrix} \rightarrow \\ \rightarrow \\ \rightarrow \\ \leftarrow \\ \rightarrow \end{bmatrix}$$



$\tau_1 = s_1, s_2, s_3, sF_2$

$\tau_2 = s_1, sF_1$

$\tau_3 = s_1, s_2, s_1, sF_1$

$\tau_4 = s_1, s_2, s_1, s_2, s_3, sF_2$

$\tau_5 = s_1, s_2, s_3, s_2, s_3, sF_2$

$\tau_6 = s_1, s_2, s_3, s_2, s_1, sF_1$

$\tau_7 = s_1, sF_1, s_1$

$\tau_8 = s_1, s_2, s_3, sF_2, s_3$

$\tau_9 = s_1, s_2, s_1, s_2, s_3, s_2, s_3, sF_2$

$\tau_{10} = s_1, s_2, s_1, s_2, s_3, s_2, s_1, sF_1$

d. Construya todas las trayectorias posibles a partir del estado s2 dada la función de acción $f_\pi(s)$ que lleven a un estado final ya sea sF1 o sF2
(NOTA: Dado que es un número infinito de trayectorias solo escriba 10)

$\tau_1 = s_2, s_3, sF_2$
$\tau_2 = s_2, s_1, sF_1$
$\tau_3 = s_2, s_3, s_2, s_1, sF_1$
$\tau_4 = s_2, s_1, s_2, s_3, sF_2$
$\tau_5 = s_2, s_1, s_2, s_3, s_2, s_1, sF_1$
$\tau_6 = s_2, s_1, sF_1, s_1$
$\tau_7 = s_2, s_3, sF_2, s_3$
$\tau_8 = s_2, s_3, s_2, s_3, sF_2$
$\tau_9 = s_2, s_3, s_2, s_1, s_2, s_3, sF_2$
$\tau_{10} = s_2, s_1, s_2, s_1, sF_1$

e. Construya todas las trayectorias posibles a partir del estado s3 dada la función de acción $f_\pi(s)$ que lleven a un estado final ya sea sF1 o sF2
(NOTA: Dado que es un número infinito de trayectorias solo escriba 10)

$\tau_1 = s_3, sF_2$
$\tau_2 = s_3, s_2, s_1, sF_1$
$\tau_3 = s_3, s_2, s_3, sF_2$
$\tau_4 = s_3, s_2, s_1, s_2, s_3, sF_2$
$\tau_5 = s_3, s_2, s_3, s_2, s_1, sF_1$
$\tau_6 = s_3, sF_2, s_3$
$\tau_7 = s_3, s_2, s_1, sF_1, s_1$
$\tau_8 = s_3, s_2, s_3, s_2, s_3, sF_2$
$\tau_9 = s_3, s_2, s_1, s_2, s_1, sF_1$
$\tau_{10} = s_3, s_2, s_3, s_2, s_3, s_2, s_3, sF_2$

d. Calcule la recompensa acumulada de cada posible trayectoria en los incisos c, d, e usando γ=0.7.

$\rightarrow s(0) = s_1$

$\tau_1 = 4.22$
$\tau_2 = -10$
$\tau_3 = -5.3$
$\tau_4 = 1.67$
$\tau_5 = 1.39$
$\tau_6 = -3.28$
$\tau_7 = -10$
$\tau_8 = 4.08$
$\tau_9 = 0.28$
$\tau_{10} = -2$

$\rightarrow s(0) = s_2$

$\tau_{11} = 6.6$
$\tau_{12} = -7$
$\tau_{13} = -4.11$
$\tau_{14} = 2.95$
$\tau_{15} = -2.29$
$\tau_{16} = -7$
$\tau_{17} = 6.4$
$\tau_{18} = 2.55$
$\tau_{19} = 0.77$
$\tau_{20} = -3.71$

$\rightarrow s(0) = s_3$

$\tau_{21} = 10$
$\tau_{22} = -5.3$
$\tau_{23} = 4.22$
$\tau_{24} = 1.67$
$\tau_{25} = -3.28$
$\tau_{26} = 9.72$
$\tau_{27} = -5.3$
$\tau_{28} = 1.39$
$\tau_{29} = -2.99$
$\tau_{30} = 2.20$

All these numbers were calculated using:

$$r = r_1 + \gamma[f_{AR}(\tau)]$$

Programmed in python

The code

```
r = [-10,0,-0.4,-0.4,10]

g = 0.7

for t in ts:
    acc = 0
    for i in range(1,len(t)):
        s = t[i]
        acc += (g**(i-1))*(r[s])
    print(f"T{ts.index(t)+1}:{acc}")
```