

Homework 02: Accumulated Reward

Mariana Ávalos Arce
0197495

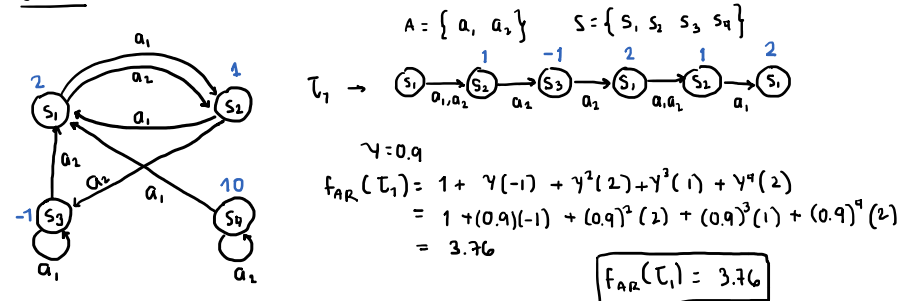
Tuesday, March 22, 2022 10:54 AM

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

$$f_{MT}(s, a) = \begin{matrix} & a_1 & a_2 \\ s_1 & s_2 & s_2 \\ s_2 & s_1 & s_3 \\ s_3 & s_3 & s_1 \\ s_4 & s_1 & s_4 \end{matrix} \quad f_R(s_f) = \begin{matrix} s_1 & 2 \\ s_2 & 1 \\ s_3 & -1 \\ s_4 & 10 \end{matrix}$$

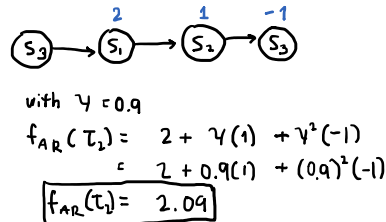
- ⓐ Calculate the accumulated reward function $f_{AR}(T_1)$ for the trajectory:
 $T_1 = s_1, s_2, s_3, s_1, s_2, s_1$

Solution



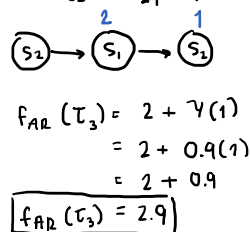
- ⓑ Calculate the accumulated reward function $f_{AR}(T_2)$ for the trajectory:
 $T_2 = s_3, s_1, s_2, s_3$

Solution



- ⓒ Calculate the accumulated reward function $f_{AR}(T_3)$ for the trajectory:
 $T_3 = s_2, s_1, s_2$

Solution

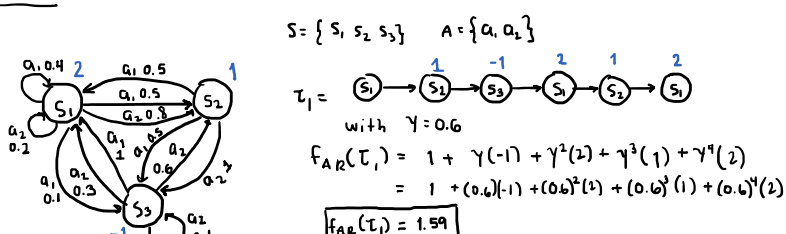


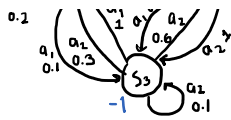
- ② Given the world defined by the transition function $P_{MT}(s_f | s, a)$, the reward function $f_R(s_f, s_t) = f_R(s_f)$ and $\gamma = 0.6$:

$$f_{MT}(s, a) = \begin{matrix} s_f = s_1 & s_f = s_2 & s_f = s_3 \\ a_1 & a_1 & a_1 \\ a_2 & a_2 & a_2 \\ s_1 & \begin{bmatrix} 0.4 & 0.2 \\ 0.5 & 0 \end{bmatrix} & \begin{bmatrix} 0.5 & 0.8 \\ 0 & 0 \end{bmatrix} & \begin{bmatrix} 0.1 & 0 \\ 0.5 & 1 \end{bmatrix} \\ s_2 & \begin{bmatrix} 0.5 & 0.8 \\ 0 & 0 \end{bmatrix} & \begin{bmatrix} 0.5 & 0.8 \\ 0 & 0 \end{bmatrix} & \begin{bmatrix} 0.5 & 1 \\ 0 & 0.1 \end{bmatrix} \\ s_3 & \begin{bmatrix} 0.5 & 0.8 \\ 0 & 0 \end{bmatrix} & \begin{bmatrix} 0.5 & 0.8 \\ 0 & 0 \end{bmatrix} & \begin{bmatrix} 0.5 & 1 \\ 0 & 0.1 \end{bmatrix} \end{matrix} \quad f_R(s_f) = \begin{matrix} s_1 & 2 \\ s_2 & 1 \\ s_3 & -1 \end{matrix}$$

- ⓐ Calculate the accumulated reward function $f_{AR}(T_1)$ for the trajectory:
 $T_1 = s_1, s_2, s_3, s_1, s_2, s_1$

Solution





$$f_{AR}(T_1) = 1 + \gamma(-1) + \gamma^2(2) + \gamma^3(1) + \gamma^4(2)$$

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

$$f_{AR}(T_1) = 1.59$$

- (b) Calculate the accumulated reward function $f_{AR}(T_2)$ for the trajectory:
 $T_2 = s_3, s_1, s_2, s_3$

Solution

with $\gamma = 0.6$

$$f_{AR}(T_2) = 2 + \gamma(1) + \gamma^2(-1)$$

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

$$f_{AR}(T_2) = 2.24$$

- (c) Calculate the accumulated reward function $f_{AR}(T_3)$ for the trajectory:
 $T_3 = s_3, s_1, s_2$

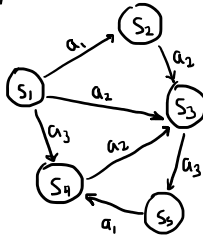
Solution

$$f_{AR}(T_3) = 2 + \gamma(1)$$

$$= 2 + 0.6(1)$$

$$f_{AR}(T_3) = 2.6$$

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



	$s_f = s_1$	$s_f = s_2$	$s_f = s_3$	$s_f = s_4$	$s_f = s_5$
	$a_1 \ a_2 \ a_3$	$a_1 \ a_2 \ a_3$	$a_1 \ a_2 \ a_3$	$a_1 \ a_2 \ a_3$	$a_1 \ a_2 \ a_3$
s_1	$\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} -2 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 5 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & -3 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$
s_2	$\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 4 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$
s_3	$\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & -6 \end{bmatrix}$
s_4	$\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & -1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$
s_5	$\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$

- (a) Calculate the accumulated reward function $f_{AR}(T_1)$ for the trajectory:
 $T_1 = s_1, s_2, s_3, s_5, s_4, s_3, s_5$

Solution

with $\gamma = 0.8$

$$f_{AR}(T_1) = -2 + \gamma(4) + \gamma^2(-6) + \gamma^3(1) + \gamma^4(-1) + \gamma^5(-6)$$

$$= -2 + (0.8)(4) + (0.8)^2(-6) + (0.8)^3(1) + (0.8)^4(-1) + (0.8)^5(-6)$$

$$f_{AR}(T_1) = -0.57$$

- (b) Calculate the accumulated reward function $f_{AR}(T_2)$ for the trajectory:
 $T_2 = s_1, s_3, s_5, s_4$

Solution

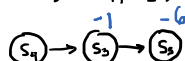
with $\gamma = 0.8$

$$f_{AR}(T_2) = 5 + \gamma(-6) + \gamma^2(1)$$

$$= 5 + (0.8)(-6) + (0.8)^2(1)$$

$$f_{AR}(T_2) = 0.84$$

- (c) Calculate the accumulated reward function $f_{AR}(T_3)$ for the trajectory:
 $T_3 = s_4, s_3, s_5$

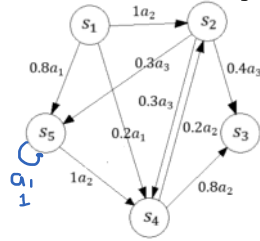


with $\gamma = 0.8$

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

$$f_{AR}(\tau_3) = 3.8$$

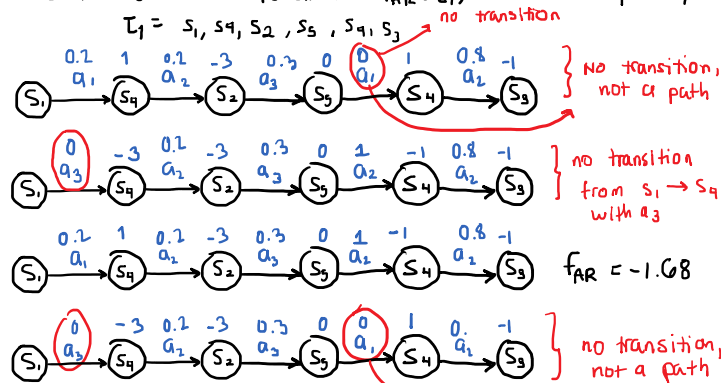
- 4) Given the world defined by the following graph where $0.8a_1$ means a_1 with probability 0.8, and so on, and with the following reward function and $\gamma = 0.7$



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

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

Solution

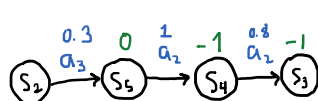
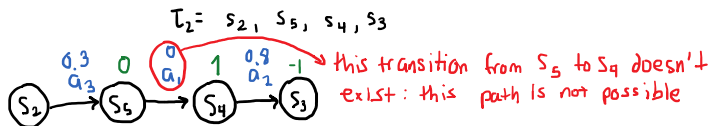


$$f_{AR} = -1.68$$

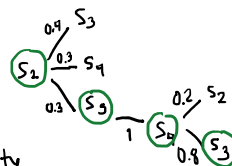
$$f_{AR} = 1 + (0.7)(-3) + (0.7)^2(0) + 0.7^3(-1) + 0.7^4(-1) = -1.68$$

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

Solution

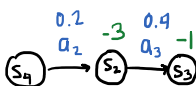
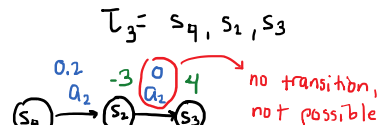


$$f_{AR} = 0 + (0.7)(-1) + (0.7)^2(-1) \\ f_{AR} = -1.19 \text{ with } 0.14 \text{ of probability}$$



- c) Calculate the accumulated reward function $f_{AR}(\tau_3)$ for the trajectory:

Solution



$$f_{AR} = -3 + (0.7)(-1) = -3.7$$

with 0.08 probability

- 5) El mundo tiene el siguiente conjunto de estados $S=\{s_1, s_2, s_3, sF_1, sF_2\}$ donde s_1 =estado inicial y, sF_1 y sF_2 son estados terminales:

sF1	s1	s2	s3	sF2
-----	----	----	----	-----

El mundo tiene el siguiente conjunto de acciones $A=\{\rightarrow, \leftarrow\}$ donde:

\rightarrow =Agente se mueve a la derecha una sola celda

\leftarrow =Agente se mueve a la izquierda una sola celda

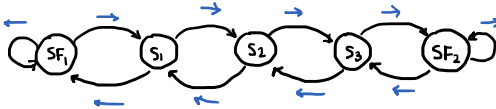
La función de recompensa $f_R(s, a, s_f) = f_R(s_f)$ solo depende del estado al que el Agente llega y esta definida como:

-10	0	-0.4	-0.4	10
-----	---	------	------	----

Es decir, si el agente transiciona de s_1 a s_2 entonces recibe la recompensa -0.4 que esta define en el estado s_2 . El agente tiene la siguiente función de acción $f_\pi(s)$:

$$f_\pi(s) = \begin{matrix} s_1 & \rightarrow \\ s_2 & \rightarrow \\ s_3 & \rightarrow \\ sF_1 & \leftarrow \\ sF_2 & \rightarrow \end{matrix}$$

- a) Build the graph of the world



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

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

- 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

$$f_\pi(s) = \begin{matrix} s_1 & \rightarrow \\ s_2 & \rightarrow \\ s_3 & \rightarrow \\ sF_1 & \leftarrow \\ sF_2 & \rightarrow \end{matrix} \quad \begin{matrix} T_1 = s_1, s_2, s_3, sF_2 \\ T_2 = s_1, s_2, s_3, sF_1, s_3, sF_2 \end{matrix}$$

- 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

$$\begin{matrix} T_1 = s_2, s_3, sF_2 \\ T_2 = s_2, s_3, sF_1, s_3, sF_2 \end{matrix}$$

- 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

$$\begin{matrix} T_1 = s_3, sF_2 \\ T_2 = s_3, sF_1, s_3, sF_2 \end{matrix}$$

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

$$\begin{matrix} s(0) = s_1 & \begin{matrix} sF_1 & s_1 & s_2 & s_3 & sF_2 \\ -10 & 0 & -0.4 & -0.4 & 10 \end{matrix} \\ T_1 = s_1, s_2, s_3, sF_2 & f_{RA} = -0.4 + (0.7)(-0.4) + (0.7)^2(10) = 4.22 \\ T_2 = s_1, s_2, s_3, sF_2, s_3, sF_2 & f_{RA} = -0.4 + (0.7)(-0.4) + (0.7)^2(10) + (0.7)^3(-0.4) + (0.7)^4(10) = 6.48 \\ s(0) = s_2 & \\ T_1 = s_2, s_3, sF_2 & f_{RA} = -0.4 + (0.7)(10) = -0.4 + 7 = 6.6 \end{matrix}$$

$$\tau_2 = s_2, s_3, s_{F1}, s_3, s_{F2}$$

$$s(0) = s_3$$

$$\tau_1 = s_3, s_{F1}$$

$$\tau_2 = s_3, s_{F1}, s_3, s_{F2}$$

$$f_{RA} = -0.4 + (0.7)(10) + (0.7)^2(0.4) + (0.7)^3(10) = 9.83$$

$$f_{RA} = -0.4 + (0.7)(10) + (0.7)^2(0.4) + (0.7)^3(10) = 9.83$$

$$f_{RA} = 10$$

$$f_{RA} = 10 + (0.7)(-0.4) + (0.7)^2(10) = 14.62$$

- 6 El mundo tiene el siguiente conjunto de estados $S = \{s_1, s_2, s_3, s_{F1}, s_{F2}\}$ donde s_1 = estado inicial y, s_{F1} y s_{F2} son estados terminales:

sF1	s1	s2	s3	sF2
-----	----	----	----	-----

El mundo tiene el siguiente conjunto de acciones $A = \{\rightarrow, \leftarrow\}$ donde:

\rightarrow = 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

\leftarrow = 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:

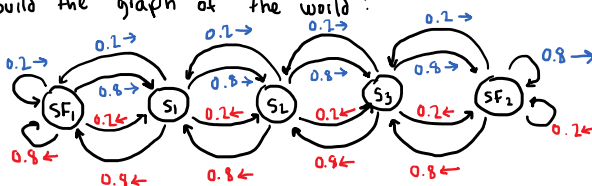
-10	0	-0.4	-0.4	10
-----	---	------	------	----

Es decir, si el agente transiciona de s_1 a s_2 entonces recibe la recompensa -0.4 que esta definide en el estado s_2 .

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

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

- a Build the graph of the world :

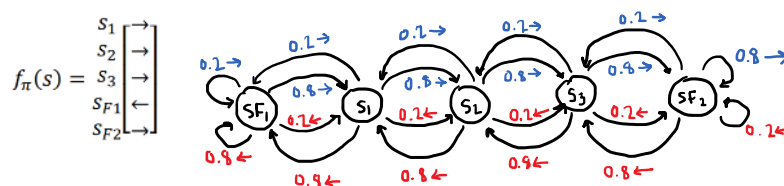


- b Write the transition function $P_{MT}(s_f | s, a)$

$$P_{MT}(s_f | s, a) = \begin{matrix} & \leftarrow & \rightarrow \\ \begin{matrix} s_1 \\ s_2 \\ s_3 \\ s_{F1} \\ s_{F2} \end{matrix} & \begin{bmatrix} 0 & 0 \\ 0.8 & 0.2 \\ 0 & 0 \\ 0.2 & 0.8 \\ 0 & 0 \end{bmatrix} & \begin{bmatrix} 0.2 & 0.8 \\ 0 & 0 \\ 0.8 & 0.2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 \\ 0.2 & 0.8 \\ 0 & 0 \\ 0.8 & 0.2 \\ 0 & 0 \end{bmatrix} & \begin{bmatrix} 0.8 & 0.2 \\ 0 & 0 \\ 0 & 0 \\ 0.8 & 0.2 \\ 0 & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0.2 & 0.8 \\ 0 & 0 \\ 0.2 & 0.8 \end{bmatrix} \end{matrix}$$

c. Construya todas las trayectorias posibles a partir del estado inicial s_1 dada la función de acción $f_\pi(s)$ que lleven a un estado final ya sea s_{F1} o s_{F2}

(NOTA: Dado que es un número infinito de trayectorias solo escriba 10)



$$\tau_1 = s_1, s_2, s_3, s_{F1}$$

$$\tau_2 = s_1, s_{F1}$$

$$\tau_3 = s_1, s_2, s_1, s_{F1}$$

$$\tau_4 = s_1, s_2, s_1, s_2, s_3, s_{F2}$$

$$\tau_5 = s_1, s_2, s_3, s_2, s_3, s_{F2}$$

$$\tau_6 = s_1, s_2, s_3, s_2, s_1, s_{F1}$$

$$\tau_7 = s_1, s_{F1}, s_1$$

$$\tau_8 = s_1, s_2, s_3, s_{F2}, s_3$$

$$\tau_9 = s_1, s_2, s_1, s_2, s_3, s_2, s_3, s_{F2}$$

$$\tau_{10} = s_1, s_2, s_1, s_2, s_3, s_2, s_1, s_{F1}$$

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)

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

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)

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

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

$\rightarrow s(0)=s_1$	$\rightarrow s(0)=s_2$
$\tau_1 = 4.22$	$\tau_{11} = 6.6$
$\tau_2 = -10$	$\tau_{12} = -7$
$\tau_3 = -5.3$	$\tau_{13} = -4.11$
$\tau_4 = 1.67$	$\tau_{14} = 2.95$
$\tau_5 = 1.39$	$\tau_{15} = -2.29$
$\tau_6 = -3.28$	$\tau_{16} = -7$
$\tau_7 = -10$	$\tau_{17} = 6.4$
$\tau_8 = 4.08$	$\tau_{18} = 2.55$
$\tau_9 = 0.28$	$\tau_{19} = 0.77$
$\tau_{10} = -2$	$\tau_{20} = -3.71$

$\rightarrow s(0)=s_3$

$\tau_{21} = 10$	$\tau_{26} = 9.72$
$\tau_{22} = -5.3$	$\tau_{27} = -5.3$
$\tau_{23} = 4.22$	$\tau_{28} = 1.39$
$\tau_{24} = 1.67$	$\tau_{29} = -2.99$
$\tau_{25} = -3.28$	$\tau_{30} = 2.10$

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}")
```