MATERIA: Aprendizaje Automático para Grandes Volúmenes de Datos
TAREA: Aprendizaje por Refuerzo

1. Dado el mundo definido por la siguiente función de transición $f_{M_T}(s,a)$, la función de recompensa $f_R(s,a,s_f) = f_R(s_f)$ y γ=0.9:

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

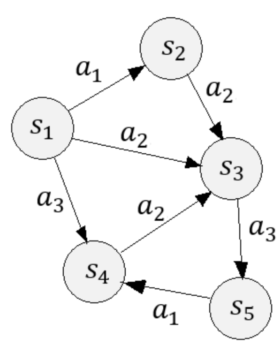
a. Calcule la función de la recompensa acumulada $f_{RA}(\tau_1)$ para la trayectoria
τ_1= s1,s2,s3,s1,s2,s1
b. Calcule la función de la recompensa acumulada $f_{RA}(\tau_2)$ para la trayectoria τ_2=s3,s1,s2,s3
c. Calcule la función de la recompensa acumulada $f_{RA}(\tau_3)$ para la trayectoria τ_3=s2,s1,s2

2. Dado el mundo definido por la función de transición $P_{M_T}(s_f|s,a)$, la función de recompensa $f_R(s,a,s_f) = f_R(s_f)$ y γ=0.6:

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

a. Calcule la función de la recompensa acumulada $f_{RA}(\tau_1)$ para la trayectoria
τ_1= s1,s2,s3,s1,s2,s1
b. Calcule la función de la recompensa acumulada $f_{RA}(\tau_2)$ para la trayectoria τ_2=s3,s1,s2,s3
c. Calcule la función de la recompensa acumulada $f_{RA}(\tau_3)$ para la trayectoria τ_3=s3,s1,s2

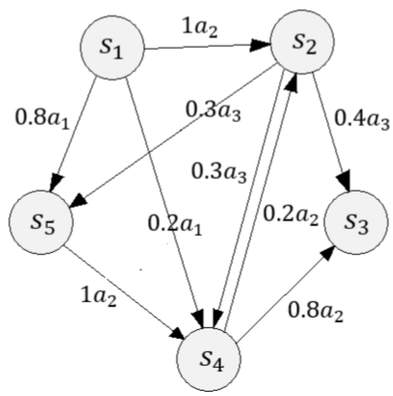
3. Dado el mundo definido por el siguiente grafo, la siguiente función de recompensa $f_R(s,a,s_f)$ y γ=0.8:



$$
f_{\mathrm{R}}(s,a,s_f) =
\begin{array}{c} s_f = s_1 \\ \begin{array}{cc} & \begin{array}{ccc} a_1 & a_2 & a_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \end{array} & \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array} \end{array}
\begin{array}{c} s_f = s_2 \\ \begin{array}{cc} & \begin{array}{ccc} a_1 & a_2 & a_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \end{array} & \begin{bmatrix} -2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array} \end{array}
\begin{array}{c} s_f = s_3 \\ \begin{array}{cc} & \begin{array}{ccc} a_1 & a_2 & a_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \end{array} & \begin{bmatrix} 0 & 5 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array} \end{array}
\begin{array}{c} s_f = s_4 \\ \begin{array}{cc} & \begin{array}{ccc} a_1 & a_2 & a_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \end{array} & \begin{bmatrix} 0 & 0 & -3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \end{array} \end{array}
\begin{array}{c} s_f = s_5 \\ \begin{array}{cc} & \begin{array}{ccc} a_1 & a_2 & a_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \end{array} & \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -6 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array} \end{array}
$$

a. Calcule la función de la recompensa acumulada $f_{\mathrm{RA}}(\tau_1)$ para la trayectoria τ_1= s1,s2,s3,s5,s4,s3,s5
b. Calcule la función de la recompensa acumulada $f_{\mathrm{RA}}(\tau_2)$ para la trayectoria τ_2=s1,s3,s5,s4
c. Calcule la función de la recompensa acumulada $f_{\mathrm{RA}}(\tau_3)$ para la trayectoria τ_3=s4,s3,s5

4. Dado el mundo definido por el siguiente grafo donde 0.8a_1 significa a_1 con probabilidad 0.8, 1a_2 significa a_2 con probabilidad 1, y así sucesivamente, y con la siguiente función de recompensa $f_{\mathrm{R}}(s,a,s_f)$ y γ=0.7:



$$
f_{\mathrm{R}}(s,a,s_f) =
\begin{array}{c} s_f = s_1 \\ \begin{array}{cc} & \begin{array}{ccc} a_1 & a_2 & a_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \end{array} & \begin{bmatrix} 9 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \\ 2 & 0 & 0 \end{bmatrix} \end{array} \end{array}
\begin{array}{c} s_f = s_2 \\ \begin{array}{cc} & \begin{array}{ccc} a_1 & a_2 & a_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \end{array} & \begin{bmatrix} -2 & 2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array} \end{array}
\begin{array}{c} s_f = s_3 \\ \begin{array}{cc} & \begin{array}{ccc} a_1 & a_2 & a_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \end{array} & \begin{bmatrix} 0 & 5 & 0 \\ 0 & 4 & -1 \\ 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array} \end{array}
\begin{array}{c} s_f = s_4 \\ \begin{array}{cc} & \begin{array}{ccc} a_1 & a_2 & a_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \end{array} & \begin{bmatrix} 1 & 0 & -3 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & -1 & 0 \end{bmatrix} \end{array} \end{array}
\begin{array}{c} s_f = s_5 \\ \begin{array}{cc} & \begin{array}{ccc} a_1 & a_2 & a_3 \end{array} \\ \begin{array}{c} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \end{array} & \begin{bmatrix} 3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -6 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array} \end{array}
$$

a. Calcule la función de la recompensa acumulada $f_{\mathrm{RA}}(\tau_1)$ para la trayectoria τ_1= s1,s4,s2,s5,s4,s3
b. Calcule la función de la recompensa acumulada $f_{\mathrm{RA}}(\tau_2)$ para la trayectoria τ_2=s2,s5,s4,s3
c. Calcule la función de la recompensa acumulada $f_{\mathrm{RA}}(\tau_3)$ para la trayectoria τ_3=s4,s2,s3

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 $f_\pi(s)$:

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

Haga lo siguiente:

a. Construya el grafo del mundo. (NOTA: Igual que en la Tarea 1)

b. Escriba la función de transición $f_{M_T}(s,a)$. (NOTA: Igual que en la Tarea 1)

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

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

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

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

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 $f_\pi(s)$:

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

Haga lo siguiente:

a. Construya el grafo del mundo. (NOTA: Igual que en la Tarea 1)

b. Escriba la función de transición $P_{M_T}(s_f|s,a)$. (NOTA: Igual que en la Tarea 1)

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)

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)

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)

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