

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 $\gamma=0.9$:

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

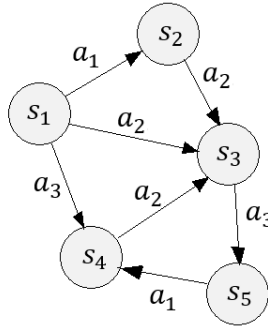
- a. Resuelva las Ecuaciones de Optimalidad de Bellman por Iteración de Valor para $V(s)$.
- b. Resuelva las Ecuaciones de Optimalidad de Bellman por Iteración de Valor para $Q(s,a)$.

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 $\gamma=0.6$:

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

- a. Resuelva las Ecuaciones de Optimalidad de Bellman por Iteración de Valor para $V(s)$.
- b. Resuelva las Ecuaciones de Optimalidad de Bellman por Iteración de Valor para $Q(s,a)$.

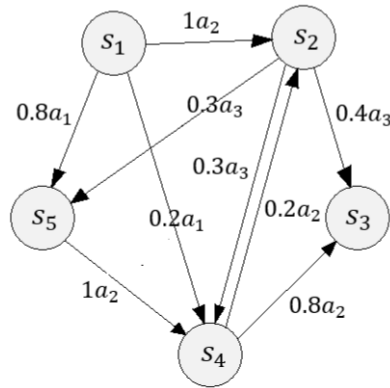
3. Dado el mundo definido por el siguiente grafo, la siguiente función de recompensa $f_R(s, a, s_f)$ y $\gamma=0.8$:



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

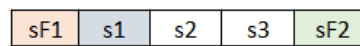
- a. Resuelva las Ecuaciones de Optimalidad de Bellman por Iteración de Valor para $V(s)$.
- b. Resuelva las Ecuaciones de Optimalidad de Bellman por Iteración de Valor para $Q(s,a)$.

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 $f_R(s, a, s_f)$ y $\gamma=0.7$:



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

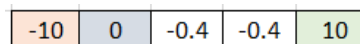
- a. Resuelva las Ecuaciones de Optimalidad de Bellman por Iteración de Valor para $V(s)$.
b. Resuelva las Ecuaciones de Optimalidad de Bellman por Iteración de Valor para $Q(s,a)$.
5. El mundo tiene el siguiente conjunto de estados $S=\{s1, s2, s3, sF1, sF2\}$ donde $s1$ =estado inicial y, $sF1$ y $sF2$ son estados terminales:



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:



Es decir, si el agente transiciona de $s1$ a $s2$ entonces recibe la recompensa -0.4 que esta define en el estado $s2$. Use $\gamma=0.8$

Haga lo siguiente:

- a. Resuelva las Ecuaciones de Optimalidad de Bellman por Iteración de Valor para $V(s)$.
b. Resuelva las Ecuaciones de Optimalidad de Bellman por Iteración de Valor para $Q(s,a)$.

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 $A=\{\rightarrow, \leftarrow\}$ 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 $s1$ a $s2$ entonces recibe la recompensa -0.4 que esta define en el estado $s2$. Use $\gamma=0.6$

Haga lo siguiente:

- a. Resuelva las Ecuaciones de Optimalidad de Bellman por Iteración de Valor para $V(s)$.
- b. Resuelva las Ecuaciones de Optimalidad de Bellman por Iteración de Valor para $Q(s,a)$.