

MATERIA: Aprendizaje Automático para Grandes Volúmenes de Datos

TAREA: Aprendizaje por Refuerzo

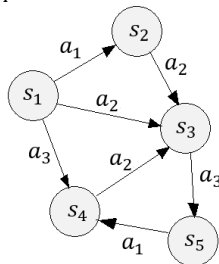
1. Construya el grafo del mundo definido por la siguiente función de transición $f_{M_T}(s, a)$:

$$f_{M_T}(s, a) = \begin{matrix} & a_1 & a_2 \\ \begin{matrix} s_1 \\ s_2 \\ s_3 \\ s_4 \end{matrix} & \begin{bmatrix} s_2 & s_2 \\ s_1 & s_3 \\ s_3 & s_1 \\ s_1 & s_4 \end{bmatrix} \end{matrix}$$

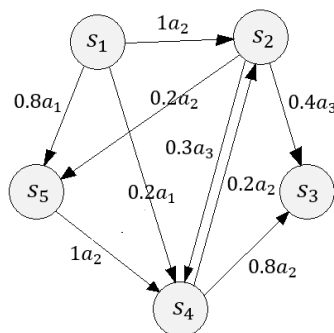
2. Construya el grafo del mundo definido por la siguiente función de transición $P_{M_T}(s_f | s, a)$:

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

3. Construya la función de transición $f_{M_T}(s, a)$ del siguiente mundo:



4. Construya la función de transición $P_{M_T}(s_f | s, a)$ del siguiente mundo:



Donde $0.8a_1$ significa a_1 con probabilidad 0.8, $1a_2$ significa a_2 con probabilidad 1, y así sucesivamente.

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 $A=\{\rightarrow, \leftarrow\}$ donde:

\rightarrow =Agente se mueve a la derecha

\leftarrow =Agente se mueve a la izquierda

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 $s2$.

Haga lo siguiente:

a. Construya el grafo del mundo

b. Escriba la función de transición $f_{M_T}(s, a)$

c. Escriba la función de recompensa $f_R(s, a, s_f) = f_R(s_f)$

d. Escriba 2 funciones de acción $f_\pi(s)$:

I. Una para que el Agente a partir del estado inicial $s1$ llegue al estado final $sF2$

II. Una para que el Agente a partir del estado inicial $s1$ llegue al estado final $sF1$

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 con probabilidad 0.8 y se mueve a la izquierda con probabilidad 0.2

\leftarrow =Agente se mueve a la izquierda con probabilidad 0.8 y se mueve 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 $s2$.

Haga lo siguiente:

- a. Construya el grafo del mundo
- b. Escriba la función de transición $P_{M_T}(s_f|s, a)$
- c. Escriba la función de recompensa $f_R(s, a, s_f) = f_R(s_f)$
- d. Escriba 2 funciones de acción $f_\pi(s)$:
 - I. Una para que el Agente a partir del estado inicial s1 llegue al estado final sF2
 - II. Una para que el Agente a partir del estado inicial s1 llegue al estado final sF1
- e. Muestre con que probabilidad el agente llega al estado final sF2 y al estado final sF1 en ambas políticas del inciso d.