

1.

1 / 1 punto

¿Cuál de las siguientes describe con precisión la función de valor de estado-acción $Q(s, a)$?

- ☒ Es la suma de la recompensa inmediata y el valor de estado $V(s)$ después de tomar la acción a (una vez), luego comportarse de manera óptima después de eso.
- ☐ Es la suma de la recompensa inmediata y el valor de estado $V(s)$ después de tomar la acción a repetidamente.
- ☐ Es la suma de la recompensa inmediata y el valor de estado $V(s)$ después de comportarse de manera óptima.
- ☐ Es la recompensa inmediata si parte de estado s y tomar la acción a (una vez).

✓ **Correcto**
¡Excelente!

2.

1 / 1 punto

Estás controlando un robot que tiene 3 acciones: \leftarrow (izquierda), \rightarrow (derecha) y STOP. De un estado s , ha calculado $Q(s, \leftarrow) = -10$, $Q(s, \rightarrow) = -20$, $Q(s, \text{STOP}) = 0$.

¿Cuál es la acción óptima a tomar en el estado s ?

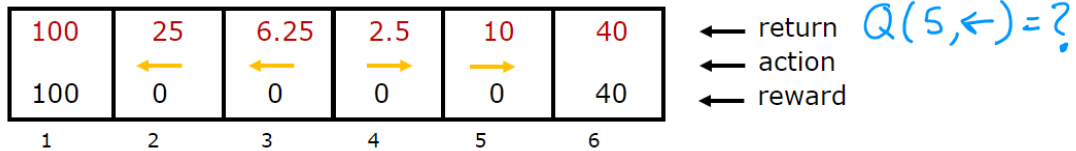
- ☒ DETÉNGASE
- ☐ \leftarrow (izquierda)
- ☐ \rightarrow (derecha)
- ☐ Imposible de decir

✓ **Correcto**
Sí, porque esto tiene el mayor valor.

3.

1 / 1 punto

Para este problema, $C = 0.25$. El siguiente diagrama muestra el retorno y la acción óptima de cada estado. Calcule $Q(5, \leftarrow)$.



☒ 0.625

☐ 0.391

☐ 1.25

☐ 2.5

☒ **Correcto**

si, lo conseguimos 0 recompensa en el estado 5. Entonces $0 * 0.25$ recompensa con descuento en el estado 4, ya que nos movimos a la izquierda para nuestra acción. Ahora nos comportamos de manera óptima a partir del estado 4 en adelante. Entonces, nos movemos a la derecha al estado 5 desde el estado 4 y recibimos $0 * 0.25^2$ recompensa con descuento. Finalmente, nos movemos a la derecha en el estado 5 al estado 6 para recibir una recompensa con descuento de $40 * 0.25^3$. Sumando estos juntos obtenemos 0.625.