

Cayetano Molina 20211
Estefanía Elvira 20725
Priscilla González Sandoval 20689

Task 1

1. ¿Qué es un Markov Decision Process (MDP)?
 - a. Una MDP se trata de un tipo de machine learning en el que se quiere aumentar el rendimiento de la máquina. Lo que hacen es tener una recompensa que se conoce como señal de refuerzo para que así el agente sea capaz de tomar decisiones en torno a un ambiente en específico o estocástico.
2. ¿Cuáles son los componentes principales de un MDP?
 - a. Los componentes principales de un MDP son:
 - Estados, acciones, función de transición, recompensa y política.
3. ¿Cuál es el objetivo principal del aprendizaje por refuerzo con MDPs?
 - a. El objetivo principal se trata acerca de que el agente explore distintas situaciones y que en base al ambiente o entorno en el que se encuentre, haga que la recompensa sea siempre mayor y mejor optimizada.

Task 3

Se sabe esto ya que como lo dice el libro, π^* se denota como la póliza óptima general, se podría decir. Entonces, todas llevan la misma función de valor-estado que se denomina como la función de valor-estado óptima.

$$v_*(s) = \max_{\pi} v_{\pi}(s),$$

$$q_*(s, a) = \max_{\pi} q_{\pi}(s, a),$$

La función V lo que significa es la recompensa máxima que se puede tener por medio de un estado y la función Q es la recompensa máxima que se puede tener por medio de la acción que se tome

Referencias

GeeksforGeeks. (2018, January 4). *Markov Decision Process*. GeeksforGeeks;

GeeksforGeeks. <https://www.geeksforgeeks.org/markov-decision-process/>

Sutton, H. (2014). Peter Morgan Sutton. *BMJ*, 348(mar 31 11), g2466–g2466.

<https://doi.org/10.1136/bmj.g2466>