

Examen 1

Aprendizaje por refuerzo

Emmanuel Peto Gutiérrez

12 de abril de 2023

Problema 1

- **¿Cuál es el objetivo de aprendizaje por refuerzo?**

Diseñar agentes capaces de mejorar con la experiencia en un problema dado.

- **Explique con sus propias palabras que es un MDP.**

Es una tarea de aprendizaje por refuerzo que satisface la propiedad de Markov.

- **¿Qué representan los estado-valor y estado acción?**

El estado-valor es la recompensa esperada por llegar a cierto estado. El estado-acción es la recompensa esperada por estar en un estado y tomar cierta acción.

- **¿Cuál es la diferencia entre explorar y explotar?**

La exploración encuentra más información acerca del ambiente y la explotación utiliza la información conocida para maximizar la recompensa.

- **Explique el principio de control de Monte-Carlos, SARSA y aprendizaje Q**

- **Monte Carlo.** El método de Monte-Carlo sólo requiere ejemplares de experiencia, los cuales son secuencias de estados, acciones y recompensas.
- **SARSA.** Se sigue el patrón de iteración generalizada de política pero usando métodos de diferencia temporal para la parte de la evaluación. Es un método dentro de política (on-policy).
- **Q-learning.** Es un método de control por diferencia temporal fuera de política (off-policy). En este caso, la función aprendida acción-valor (Q) aproxima directamente Q^* .