Universidad del Valle de Guatemala
CC3085 – Inteligencia Artificial
Sección 10
24 de Marzo de 2025

Nombres: Daniela Navas **Carné:** 211000

LABORATORIO 7

Task 1 - Preguntas Teóricas

Responda a cada de las siguientes preguntas de forma clara y lo más completamente posible.

1. ¿Qué es el temporal difference learning y en qué se diferencia de los métodos tradicionales de aprendizaje supervisado? Explique el concepto de "error de diferencia temporal" y su papel en los algoritmos de aprendizaje por refuerzo

TD Learning es un método de aprendizaje por refuerzo que combina ideas del aprendizaje supervisado y no supervisado. A diferencia de los métodos tradicionales de aprendizaje supervisado, que requieren un conjunto de datos etiquetados, TD Learning utiliza la experiencia directa del agente para aprender. El **"error de diferencia temporal"** es la diferencia entre el valor predicho y el valor real observado en el siguiente estado. Este error se utiliza para ajustar las estimaciones futuras. En los algoritmos de aprendizaje por refuerzo, el error de diferencia temporal ayuda a actualizar las políticas y valores de estado de manera más eficiente.

2. En el contexto de los juegos simultáneos, ¿cómo toman decisiones los jugadores sin conocer las acciones de sus oponentes? De un ejemplo de un escenario del mundo real que pueda modelarse como un juego simultáneo y discuta las estrategias que los jugadores podrían emplear en tal situación

En juegos simultáneos, los jugadores toman decisiones sin conocer las acciones de sus oponentes. Esto se hace mediante la evaluación de estrategias y posibles resultados. Un ejemplo podría ser una subasta silenciosa, donde los participantes ofrecen precios sin saber las ofertas de los demás. Las estrategias pueden incluir ofrecer un precio que maximice la probabilidad de ganar sin exceder el valor percibido del objeto.

3. ¿Qué distingue los juegos de suma cero de los juegos de suma no cero y cómo afecta esta diferencia al proceso de toma de decisiones de los jugadores? Proporcione al menos un ejemplo de juegos que entren en la categoría de juegos de no suma cero y discuta las consideraciones estratégicas únicas involucradas

En los juegos de suma cero, la ganancia de un jugador es exactamente igual a la pérdida del otro. En los juegos de no suma cero, los jugadores pueden beneficiarse mutuamente. Como ejemplo esta la negociación comercial entre dos empresas. Aquí, ambas partes pueden encontrar una solución que beneficie a ambas, como un acuerdo de colaboración.

4. ¿Cómo se aplica el concepto de equilibrio de Nash a los juegos simultáneos? Explicar cómo el equilibrio de Nash representa una solución estable en la que ningún jugador tiene un incentivo para desviarse unilateralmente de la estrategia elegida

El equilibrio de Nash se aplica cuando cada jugador elige una estrategia óptima, considerando las estrategias de los demás. Ningún jugador tiene un incentivo para cambiar unilateralmente su

estrategia. La **estabilidad** representa una solución estable porque todos los jugadores están en su mejor respuesta dada la estrategia de los otros jugadores.

5. Discuta la aplicación del temporal difference learning en el modelado y optimización de procesos de toma de decisiones en entornos dinámicos. ¿Cómo maneja el temporal difference learning el equilibrio entre exploración y explotación y cuáles son algunos de los desafíos asociados con su implementación en la práctica?

- **Modelado y optimización:** TD Learning se utiliza para modelar y optimizar procesos de toma de decisiones en entornos donde las condiciones cambian con el tiempo.
- **Equilibrio entre exploración y explotación:** TD Learning maneja este equilibrio mediante la actualización continua de las estimaciones de valor, permitiendo al agente explorar nuevas estrategias mientras explota las conocidas.
- **Desafíos:** Algunos desafíos incluyen la necesidad de un balance adecuado entre exploración y explotación y la complejidad computacional asociada con la actualización de valores en tiempo real.

Task 2 – Connect Four

Ahora, haga que el agente entrenado con TD learning, juegue contra el agente que usa Minimax, y luego contra el agente de minimax con poda alpha-beta. Haga que estos 3 tipos de juegos sucedan por lo menos 50 veces cada uno, es decir 150 juegos en total. Con el resultado de estos 150 juegos, grafique la cantidad de victorias de cada uno de los agentes y coloque las en un documento PDF que deberá subir junto con su código en la entrega.

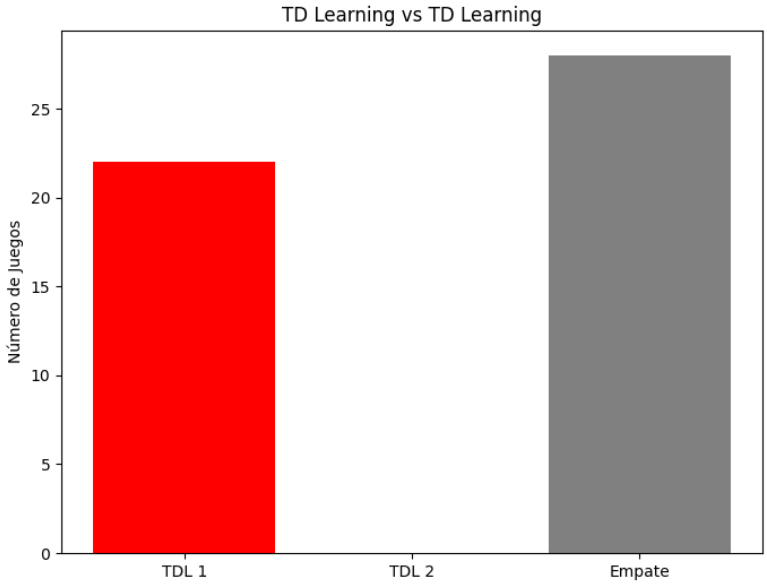

Figura 1. Resultados de Juegos | TD Learning vs TD Learning

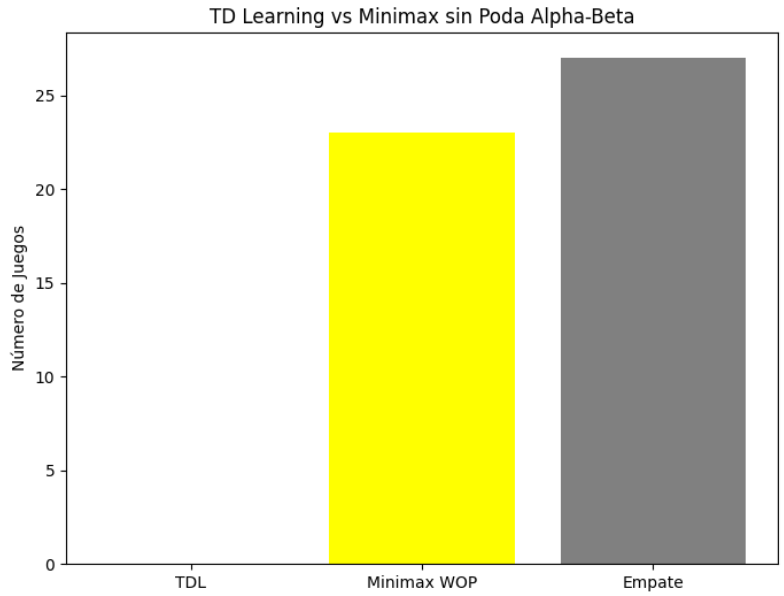

Figura 1. Resultados de Juegos | TD Learning vs Minimax sin Poda Alpha-Beta

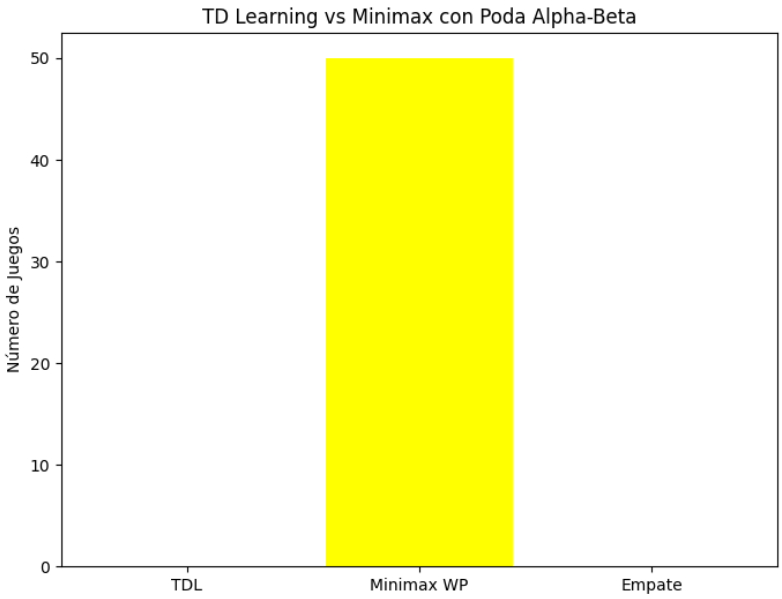

Figura 1. Resultados de Juegos | TD Learning vs Minimax con Poda Alpha-Beta

Deberá grabar un video, en el cual deberán mostrar solamente 3 juegos, es decir, uno de cada caso. Para todos los juegos en el video, asegúrense de acelerar lo suficiente para que el video no tome más de 10 minutos en total. En dicho video, también deberá mencionar (siempre dentro del marco de los 10 minutos de tiempo):

- Qué hace su agente entrenando con TD learning a nivel general
- Explique por qué ganó más veces el agente que ganó. ¿Cómo afectó el tener o no esta estrategia al agente que ganó?

GITHUB:
https://github.com/danielanavas2002/InteligenciaArtificial/tree/main/Laboratorio/Laboratorio07

YOUTUBE:
https://youtu.be/qqGaqo-NtTM