

Hoja de Trabajo 2

María Marta Ramirez Gil 21342
21438

Gustavo Andrés Gonzalez Pineda

Universidad del Valle de Guatemala

Inteligencia Artificial

Task 1 - Preguntas Teóricas

1. **Defina el proceso de decisión de Markov (MDP) y explique sus componentes.**

El Proceso de Decisión de Markov (MDP) es un marco matemático utilizado en la teoría de la toma de decisiones para modelar situaciones en las que se deben tomar decisiones secuenciales en un entorno estocástico. Sus componentes principales son:

- **Espacio de estados (S):** Representa todos los posibles estados en los que puede encontrarse el sistema en un momento dado.
 - **Espacio de acciones (A):** Define las acciones disponibles que un agente puede tomar en cada estado.
 - **Función de transición de estados (T):** Especifica la probabilidad de transición de un estado a otro dado una acción.
 - **Función de recompensa (R):** Asigna una recompensa numérica a cada transición de estado-acción, que indica la utilidad o el valor asociado con esa transición.
-
2. **Describa cual es la diferencia entre política, evaluación de políticas, mejora de políticas e iteración de políticas en el contexto de los PDM.**
- **Política:** Es una estrategia o plan que el agente sigue para tomar decisiones. Puede ser estocástica (probabilística) o determinista.

- **Evaluación de políticas:** Consiste en calcular el valor esperado de una política dada, es decir, cuánto valor acumulará el agente siguiendo esa política.
- **Mejora de políticas:** Implica encontrar una nueva política que sea mejor o al menos igual que la política actual, generalmente mediante la modificación de la política o cambiando las acciones en ciertos estados.
- **Iteración de políticas:** Es un proceso iterativo que combina la evaluación y la mejora de políticas hasta que se encuentra una política óptima o cercana a la óptima.

3. **Explique el concepto de factor de descuento (gamma) en los MDP. ¿Cómo influye en la toma de decisiones?**

El factor de descuento (gamma) en los MDP es un parámetro que controla la importancia relativa de las recompensas a corto plazo versus las recompensas a largo plazo. Un valor de gamma cercano a 1 indica que el agente valora más las recompensas futuras, mientras que un valor cercano a 0 da más peso a las recompensas inmediatas. Influye en la toma de decisiones al permitir que el agente considere el valor futuro de las acciones, lo que puede conducir a una planificación a largo plazo y a una toma de decisiones más estratégica.

4. **Analice la diferencia entre los algoritmos de iteración de valores y de iteración de políticas para resolver MDP.**

La diferencia entre los algoritmos de iteración de valores y de iteración de políticas radica en su enfoque para encontrar la política óptima:

- **Iteración de valores:** Comienza evaluando una política y luego mejora iterativamente los valores de los estados hasta que convergen a la función de valor óptima. Luego, se extrae la política óptima de estos valores.
- **Iteración de políticas:** Comienza con una política inicial y mejora directamente la política en cada iteración hasta que converge a la política óptima. No calcula explícitamente la función de valor.

5. **¿Cuáles son algunos desafíos o limitaciones comunes asociados con la resolución de MDP a gran escala? Discuta los enfoques potenciales para abordar estos desafíos.**

- **Explosión combinatoria:** A medida que aumenta el número de estados y acciones, el espacio de búsqueda crece exponencialmente, lo que dificulta encontrar soluciones óptimas.
- **Requerimientos computacionales elevados:** La evaluación y mejora de políticas pueden ser computacionalmente intensivas, especialmente en entornos complejos.
- **Problemas de convergencia:** Algunos algoritmos pueden converger lentamente o quedar atrapados en óptimos locales en lugar de globales.
- **Manejo de información incompleta o incierta:** En entornos donde la información es parcial o incierta, el agente puede enfrentar dificultades para tomar decisiones óptimas.

Para abordar estos desafíos, se pueden emplear enfoques como el uso de técnicas de aproximación de funciones, técnicas de muestreo para estimar la función de valor, paralelización de algoritmos para reducir el tiempo de ejecución, y métodos de aprendizaje por refuerzo profundo que pueden manejar espacios de estado y acción de alta dimensionalidad. Además, el diseño cuidadoso del modelo de MDP y la selección de algoritmos adecuados pueden ayudar a mitigar estas limitaciones.

Task 2 - Preguntas Analíticas

1. **Analice críticamente los supuestos subyacentes a la propiedad de Markov en los Procesos de Decisión de Markov (MDP). Analice escenarios en los que estos supuestos puedan no ser válidos y sus implicaciones para la toma de decisiones.**

La propiedad de Markov en los MDP implica que el estado futuro de un sistema depende únicamente del estado presente y de la acción tomada, sin

importar el historial de estados y acciones anteriores. Sin embargo, este supuesto puede no ser válido en varios escenarios:

- **Dependencia de estado no Markoviana:** En situaciones donde el estado futuro depende de eventos pasados que no están incluidos en el estado presente, la propiedad de Markov se viola. Por ejemplo, en juegos de azar donde el historial de resultados afecta las probabilidades futuras.
- **Observaciones parciales:** Cuando el agente no tiene acceso completo al estado del sistema, sino solo a una parte de la información, la propiedad de Markov puede no ser válida. Por ejemplo, en entornos donde ciertas variables relevantes están ocultas o son desconocidas para el agente.
- **Cambios en el entorno:** Si el entorno es dinámico y cambia con el tiempo de una manera no predecible o no Markoviana, la propiedad de Markov puede ser inválida. Por ejemplo, en sistemas donde los parámetros del entorno evolucionan de manera impredecible o son afectados por factores externos.

Las implicaciones de estos supuestos inválidos para la toma de decisiones pueden ser significativas. Si el agente basa sus decisiones en la propiedad de Markov cuando no es válida, podría subestimar la incertidumbre o tomar decisiones subóptimas debido a una comprensión incompleta del entorno.

2. Explore los desafíos de modelar la incertidumbre en los procesos de decisión de Markov (MDP) y analice estrategias para una toma de decisiones sólida en entornos inciertos.

Modelar la incertidumbre en MDP es fundamental para la toma de decisiones precisa, pero también presenta ciertos desafíos:

- **Incetidumbre estocástica:** La incertidumbre en las transiciones de estado y las recompensas puede dificultar la predicción de los resultados de las acciones.
- **Observaciones parciales o incompletas:** Si el agente no puede observar directamente el estado completo del sistema, debe lidiar con la incertidumbre asociada con las observaciones parciales.

- **Cambios en el entorno:** La incertidumbre puede surgir debido a cambios imprevistos en el entorno, como la aparición de nuevos estados o acciones.

Para abordar estos desafíos y tomar decisiones sólidas en entornos inciertos, se pueden emplear varias estrategias:

- **Modelado de incertidumbre:** Utilizar técnicas como la modelización de transiciones estocásticas y la estimación de distribuciones de probabilidad para capturar la incertidumbre en las transiciones de estado y las recompensas.
- **Exploración activa:** Incorporar la exploración activa en las políticas de decisión para adquirir información adicional sobre el entorno y reducir la incertidumbre.
- **Aprendizaje adaptativo:** Implementar algoritmos de aprendizaje que puedan adaptarse dinámicamente a la incertidumbre cambiante del entorno y ajustar sus estrategias de decisión en consecuencia.
- **Planificación robusta:** Diseñar políticas de decisión robustas que puedan funcionar bien incluso en presencia de incertidumbre, mediante la consideración de múltiples escenarios posibles y la selección de acciones que minimicen el riesgo esperado.

Estas estrategias pueden ayudar a mitigar los desafíos asociados con la incertidumbre en los MDP y permitir una toma de decisiones más sólida en entornos inciertos.