UNIVERSITAT DE BARCELONA

Maite López Sánchez

Autocuestionario sobre los contenidos de Aprendizaje (curso IA)

Si has entendido los contenidos relacionados con los Procesos de Decisión de Márkov deberías poder contestar las siguientes preguntas sobre ellos:

- ¿Representan problemas deterministas?
- ¿Cuáles son los componentes que definen un PDM?
- ¿Qué representa la función de transición?
- ¿Qué diferencia hay entre la función de transición y la función sucesor de los problemas de búsqueda vistos anteriormente?
- Especifica, para el ejemplo de mundo rejilla, T((1,3),right,(2,3)), T((1,3),up,(2,3)), T((1,3),up,(1,3)) y T((1,1),up,(1,1)).
- ¿Qué representa la función de recompensa?
- ¿Se pueden obtener recompensas en estados no finales?
- ¿De qué depende la utilidad que pueda obtener un agente?
- ¿Qué es la hipótesis de Markov?
- En los PDM: ¿es lo mismo la probabilidad de un estado sucesor condicionada a un estado actual y una acción que la probabilidad de ese mismo estado sucesor condicionada, no únicamente al estado actual y a la acción, sino también a todos los estados y acciones previos a dicho estado actual"?
- ¿Qué es la solución de un PDM?
- ¿Qué es una política?
- ¿Es lo mismo una política que un plan?
- ¿Podemos considerar un plan como una solución de un PDM?
- ¿Qué es una política óptima?
- ¿Qué significa aprender en un PDM?
- La política óptima ¿Depende de la recompensa?
- ¿Cómo podemos especificar los estados en el ejemplo de Superior/inferior?
- ¿Hay alguna acción que no se pueda hacer en alguno de los estados del ejemplo de Superior/inferior?

- ¿Cómo se definen los estados sucesores en el ejemplo el ejemplo de Superior/inferior?
- Si haciendo una acción en un estado s únicamente tenemos dos posibles estados sucesores s' y s'' y sabemos que T(s,a,s')=p, ¿Cuánto retorna T(s,a,s'')?
- Si desde s haciendo a es imposible llegar a un estado s', ¿Cuánto retorna T(s,a,s')?

Si ya tienes claras todas estas respuestas, significa que conoces los fundamentos básicos de los PDM. Continuemos pues con preguntas relacionadas con su resolución:

- ¿Cómo podemos acumular recompensas si las secuencias de acciones pueden ser infinitas?
- ¿Qué significa un gamma pequeño?
- ¿Qué dicho popular se puede relacionar con el factor de descuento?
- ¿Qué es V*(s)?
- ¿Qué es Q*(s,a)?
- ¿Qué devuelve π*(s)?
- Siguiendo las ecuaciones de Bellman, ¿Cómo se calcula V*(s)?
- Siguiendo las ecuaciones de Bellman, ¿Cómo se calcula Q*(s,a)?
- ¿Cómo calculamos la acción óptima para un estado?
- ¿Qué es la actualización de Bellman?
- ¿Cómo se inicializan los valores de los estados en el algoritmo de iteración de valores?
- ¿Cómo se aproximan los valores de los estados en el algoritmo de iteración de valores?
- ¿Es lo mismo utilidad que valor de un estado?
- ¿Cómo es el algoritmo de iteración de políticas?

Las ecuaciones de Bellman son el concepto clave, si has podido contestar las preguntas anteriores puedes ya pasar a contestar las siguientes preguntas sobre Reinforcement Learning:

- ¿Cuál es el papel de la recompensa?

- ¿Dónde realiza las acciones el agente?
- ¿Cuál es el efecto de una acción?
- ¿Cuál es el objetivo de aprender a actuar?
- ¿Hay alguna otra disciplina que estudie el aprendizaje por refuerzo?
- ¿Todas las recompensas que reciben los humanos son iguales?
- ¿Qué neurotransmisor encontramos en nuestro cerebro que nos ayuda a aprender?
- ¿Cuál es la diferencia entre el modelo de los Procesos de Decisión de Márkov y el que se usa en el aprendizaje por refuerzo?
- ¿Qué hacemos para aprender en un RL pasivo?
- ¿Cómo estimamos directamente la utilidad de un estado en el RL pasivo?
- ¿En qué consiste el aprendizaje basado en modelo?
- ¿Es el aprendizaje basado en modelo un método pasivo de aprendizaje?
- Si estando en el estado s hemos hecho a 6 veces, de las cuales 3 hemos ido a parar a s' y 2 a s'' y 1 a s''', ¿Cuáles son los valores de T(s,a,s'), T(s,a,s'') y T(s,a,s''')?
- ¿En qué consiste el algoritmo de Diferencias Temporales?
- ¿Qué diferencia hay entre el método de iteración de valores y el de Q-learning?
- ¿Qué valores actualiza de forma iterativa el método de Q-learning?
- ¿Qué es alfa en el método de Q-learning?
- ¿Qué pasa cuando usamos un valor de alfa constante en el método de Q-learning?
- ¿Cómo se tiene que definir alfa en el método de Q-learning para que calcule la media?
- ¿Cuál es la fórmula de actualización de los Q-valores que se aplica en el algoritmo de Q-learning?
- ¿Cómo se pasa de s a s'?
- ¿Qué relación hay entre el s' de una iteración y el s de la siguiente?
- El Q-learning ¿converge a una política óptima?
- ¿En qué consiste la selección de acciones épsilon-voraz?
- ¿Conviene reducir la épsilon conforme avanza el aprendizaje?