Relación 3 - Problemas de Decisión de Markov

Cuestiones

Cuestión 1. Para cada una de las afirmaciones siguientes, debes decidir si es **VERDADERA** o **FALSA**. Si crees que la afirmación es verdadera, debes dar razones que apoyen tu decisión. Si crees que es falsa, debes dar un ejemplo concreto donde no se cumpla. Las afirmaciones son las siguientes:

(a) En un MDP, si todas las recompensas son cero, entonces la valoración de un estado respecto a cualquier política es cero.

(b) En un MDP, si el factor descuento es cero, entonces la valoración de un estado es igual a su recompensa.

Problemas

Problema 1. Supongamos un proceso de decisión de Markov con estados $S = \{s_1, s_2, s_3\}$, con tres acciones posibles a_1, a_2 y a_3, tal que $A(s_1) = \{a_2, a_3\}$, $A(s_2) = \{a_1, a_3\}$ y $A(s_3) = \{a_1, a_2\}$ y con el siguiente modelo de transición (donde en cada caso, la terna de probabilidades (p_1, p_2, p_3) son las probabilidades de pasar a los estados s_1, s_2 y s_3, respectivamente):

- Aplicando a_2 a s_1: $(0, 0{,}3, 0{,}7)$.
- Aplicando a_3 a s_1: $(0{,}1, 0{,}1, 0{,}8)$.
- Aplicando a_1 a s_2: $(0{,}4, 0{,}1, 0{,}5)$.
- Aplicando a_3 a s_2: $(0{,}8, 0{,}2, 0)$.
- Aplicando a_1 a s_3: $(0{,}5, 0, 0{,}5)$.
- Aplicando a_2 a s_3: $(0, 0{,}5, 0{,}5)$.

Las recompensas son $R(s_1) = -1$, $R(s_2) = -0{,}04$ y $R(s_3) = 1$ y el factor de descuento 0.9.

Considerar la siguiente política π: $\pi(s_1) = a_3$, $\pi(s_2) = a_3$ y $\pi(s_3) = a_2$.

Se pide:

(a) ¿Cuál es la probabilidad de que se produzca la secuencia de estados s_3, s_3, s_3, s_2, s_2 aplicando la política π? ¿Qué valoración tiene esa secuencia?

(b) Plantear el sistema de ecuaciones que define V^π

(c) Plantear las ecuaciones de Bellman que definen V

(d) Supongamos que hemos resuelto las ecuaciones anteriores y que conocemos V. Describir cómo podríamos obtener la política óptima.

Problema 2. Considerando el proceso de decisión de Markov del ejemplo del movimiento del robot que se describe en las diapositivas del tema, considerando que π es la política (a) que se muestra, y con descuento 0.8, se pide:

(a) Dar la ecuación que define V^π para el estado correspondiente a la casilla $(2, 3)$.

(b) Dar la ecuación que define V para el mismo estado.

Problema 3. A lo largo de su vida, una empresa pasa por situaciones muy distintas, que por simplificar resumiremos en que al inicio de cada campaña puede estar rica o pobre, y ser conocida o desconocida. Para ello puede decidir en cada momento o bien invertir en publicidad, o bien optar por no hacer publicidad. Estas dos acciones no tienen siempre un resultado fijo, aunque podemos describirlo de manera probabilística:

(a) Si la empresa es rica y conocida y no invierte en publicidad, seguirá rica, pero existe un 50 % de probabilidad de que se vuelva desconocida. Si gasta en publicidad, con toda seguridad seguirá conocida pero pasará a ser pobre.

(b) Si la empresa es rica y desconocida y no gasta en publicidad, seguirá desconocida, y existe un 50 % de que se vuelva pobre. Si gasta en publicidad, se volverá pobre, pero existe un 50 % de probabilidades de que se vuelva conocida.

(c) Si la empresa es pobre y conocida y no invierte en publicidad, pasará a ser pobre y desconocida con un 50 % de probabilidad, y rica y conocida en caso contrario. Si gasta en publicidad, con toda seguridad seguirá en la misma situación.

(d) Si la empresa es pobre y desconocida, y no invierte en publicidad, seguirá en la misma situación con toda seguridad. Si gasta en publicidad, seguirá pobre, pero con un 50 % de posibilidades pasará aser conocida.

Supondremos que la recompensa en una campaña en la que la empresa es rica es de 10, y de 0 en en las que sea pobre. El objetivo es conseguir la mayor recompesa acumulada a lo largo del tiempo, aunque penalizaremos las ganancias obtenidas en campañas muy lejanas en el tiempo, introduciendo un factor de descuento de 0.9.

- Representar lo anterior como un proceso de decisión de Markov
- Si π es la política que consiste en invertir siempre en publicidad, plantear y resolver las ecuaciones que definen V^π
- Plantear las ecuaciones de Bellman
- Si aplicamos el algoritmo de iteración de políticas, comenzando con la política $\pi_0 = \pi$, calcular π_1 (la política que resulta tras aplicar una iteración del algoritmo).

Problema 4. Supongamos un proceso de decisión de Markov con los siguientes componentes

- Estados: s_1, s_2, s_3, s_4.
- Acciones: a_1, a_2, a_3.

- Acciones aplicables: $A(s_1) = \{a_1, a_2\}$, $A(s_2) = \{a_3\}$, $A(s_3) = \{a_2, a_3\}$ y $A(s_4) = \{a_1\}$.
- Transiciones:
  - Aplicando a_1 a s_1, se puede pasar a s_3 con probabilidad 0.2, y a s_4 con probabilidad 0.8. Aplicando a_2 a s_1, puede pasar a s_1, s_2 y s_3, con igual probabilidad.
  - Aplicando a_3 a s_2 pasamos a s_1 con toda seguridad.
  - Aplicando a_2 a s_3 puede pasar a s_3 con probabilidad $1/3$ o a s_4 con probabilidad $2/3$. Si se aplica a_3 a s_3, existe $1/3$ de probabilidad de pasar a cada uno de los restantes estados.
  - Aplicando a_1 a s_4 puede pasar a s_3, o a s_4, con igual probabilidad.
- Recompensa: $R(s_1) = -3$, $R(s_2) = -2$, $R(s_3) = 1$, $R(s_4) = 1$
- Descuento: 0,5

Responder las siguientes cuestiones:

- Considerar que π_1 es la política que aplica a_1 a s_1 y a_2 a s_3. Calcular la valoración que tiene cada estado respecto de esa política.
- Considerando la valoracion inicial V_0 tal que $V_0(s_1) = -2$, $V_0(s_2) = -1$, $V_0(s_3) = 1$ y $V_0(s_4) = 2$, aplicar una iteración del algoritmo de iteración de valores.
- Supongamos que aplicando el algoritmo anterior obtenemos finalmente una valoración V. Describir cómo se obtendría la política optima a partir de dicha valoración.

Problema 5. Supongamos un proceso de deción de Markov con los siguientes componentes

- Estados: s_1, s_2, s_3, s_4.
- Acciones: En s_2, las acciones aplicables son a_1 y a_2. En el resto, la única acción aplicable es a_1.
- Transiciones:
  - Aplicando a_1 a s_1 solo se puede pasar a s_2 o a s_3, con igual probabilidad.
  - Aplicando a_1 a s_2 pasamos a s_3 con toda seguridad.
  - Aplicando a_2 a s_2 solo se puede pasar a s_4 o permanecer en s_2, con igual probabilidad
  - Aplicando a_1 a s_3 pasamos a s_1 con toda seguridad.
  - Aplicando a_1 a s_4, nos quedamos en s_4
- Recompensa: $R(s_1) = 1$, $R(s_2) = 2$, $R(s_3) = 3$, $R(s_4) = 10$
- Descuento: 0,9

Responder las siguientes cuestiones:

- ¿Cuántas posibles políticas existen en este problema?
- Considerar que π es la política que siempre aplica a_1. Plantear las ecuaciones que definen la valoración que tiene cada estado respecto de esa política. ¿Cuál es la valoración de s_4 respecto de π?
- Plantear las ecuaciones de Bellman para este problema.
- Aplicar el algoritmo de *iteración de valores* hasta dar el valor calculado para el estado s_2 tras la **segunda iteración**. Considerar inicialmente que la valoración de cada estado es cero.

Problema 6. Supongamos un proceso de decisión de Markov con los siguientes componentes

- Estados: s_1, s_2, s_3, s_4.
- Acciones: a_1, a_2, a_3.
- Acciones aplicables: $A(s_1) = \{a_1, a_2\}$, $A(s_2) = \{a_1, a_3\}$, $A(s_3) = \{a_3\}$ y $A(s_4) = \{a_2\}$.
- Transiciones:
  - Aplicando a_1 a s_1 solo se puede pasar a s_2. Aplicando a_2 a s_1 puede pasar a s_1, s_2 y s_3, con igual probabilidad.
  - Aplicando a_1 a s_2 pasamos a s_1 con toda seguridad. Aplicando a_3 a s_2 puede pasar a s_1 con probabilidad 0,75 y a s_4 con probabilidad 0,25.
  - Aplicando a_3 a s_3 puede pasar a s_1 o a s_2 con igual probabilidad.
  - Aplicando a_2 a s_4 puede pasar a s_1, a s_2 o a s_3, con igual probabilidad.
- Recompensa: $R(s_1) = 3$, $R(s_2) = 0$, $R(s_3) = 0$, $R(s_4) = 2$
- Descuento: 0,5

Responder las siguientes cuestiones:

- ¿Cuántas posibles políticas existen en este problema?
- Considerar que π_1 es la política que siempre aplica a_1 a s_1 y a s_2. Plantear y resolver las ecuaciones que definen la valoración que tiene cada estado respecto de esa política.
- Plantear las ecuaciones de Bellman para este problema.
- Calcular cuál sería la accion asociada a s_1 tras **una** iteración del algoritmo de *iteración de políticas*, considerando que la política de inicio es π_1.