Ejercicios 8: Procesos de decisión de Markov

Departamento de Informática / Department of Computer Science
Universidad Carlos III de Madrid

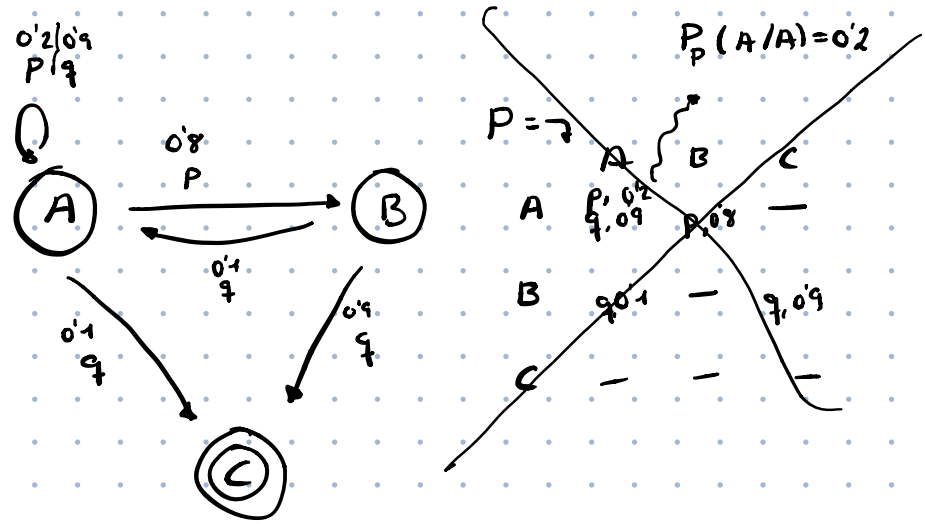
Inteligencia Artificial
Grado en Ingeniería Informática
2019/20

Ejercicio 1: MDP con tres estados

- Tenemos un agente con tres estados A, B y C donde C es el estado meta. En el estado A el agente puede llevar a cabo dos posibles acciones p y q, y en el estado B el agente puede tomar la acción q.
  - La ejecución de la acción p en el estado A mueve al agente al estado B con probabilidad 0.8, y permanece en el estado A con probabilidad 0.2.
  - La ejecución de la acción q en el estado A mueve al agente al estado C con probabilidad 0.1, y permanece en el estado A con probabilidad 0.9.
  - La ejecución de la acción q en el estado B mueve al agente al estado C con probabilidad 0.9, y mueve el agente al estado A el resto de las veces. Cada acción tiene un coste asociado de 1.
- Modelar formalmente el MDP.
- Especificar las ecuaciones de Bellman que actualizan los valores de los estados V(A) y V(B).
- Calcular el valor esperado $V(s)$ para cada estado.
- Calcular la política óptima.

Ejercicio 1)

Modelar MDP



Tupla = $\langle S, A, P, C \rangle$

$S = \{A, B, C\}$

$A = \{p, q\}$

$C = \{C(p) = 1, C(q) = 1\}$

Ecuaciones Bellman para V(A) y V(B)

$$V_{i+1}(A) = \min\left[c(p) + P_p(A/A)\, V_i(A) + P_p(B/A)\, V_i(B),\right.$$
$$\left. c(q) + P_q(C/A) V_i(C) + P_q(A/A) V_i(A)\right]$$

$$V_{i+1}(B) = \min\left[c(q) + P_q(C/B) V_i(C) + P_q(A/B) V_i(A)\right]$$

(coste de acción actual + Camino posible desde B con q · coste desde C + Camino posible desde B con q · coste desde A)

Calcular el valor esperado V(s) para cada estado

Comenzavan con valor 0: $V_0(A) = 0 \quad V_0(B) = 0 \quad V_0(C) = 0$

Paso 1: $V_1(A) = \min[1 + 0'2 \cdot 0 + 0'8 \cdot 0,\ 1 + 0'1 \cdot 0 + 0'9 \cdot 0] = 1$

$V_1(B) = \min[1 + 0'2 \cdot 0 + 0'1 \cdot 0] = 1$

$V_1(C) = 0$

Paso 2: $V_2(A) = \min[1 + 0'2 \cdot 1 + 0'8 \cdot 1,\ 1 + 0'1 \cdot 0 + 0'9 \cdot 1] = 1'9$

$V_2(B) = \min[1 + 0'1 \cdot 0 + 0'1 \cdot 1] = 1'1$

Paso 3: $V_3(A) = \min[1 + 0'2 \cdot 1'9 + 0'8 \cdot 1'1,\ 1 + 0'1 \cdot 0 + 0'9 \cdot 1'9] = 2'26$

$V_3(B) = \min[1 + 0'1 \cdot 0 + 0'1 \cdot 1'9] = 1'19$

Paso 4: $V_4(A) = \min[1 + 0'2 \cdot 2'26 + 0'8 \cdot 1'9,\ 1 + 0'1 \cdot 0 + 0'9 \cdot 2'26] = 2'4$

$V_4(B) = \min[1 + 0'1 \cdot 0 + 0'1 \cdot 2'26] = 1'2$

Paso 5: $V_5(A) = \min[1 + 0'2 \cdot 2'4 + 0'8 \cdot 1'2,\ 1 + 0'1 \cdot 0 + 0'9 \cdot 2'4] = 2'44$

$V_5(B) = \min[1 + 0'1 \cdot 0 + 0'1 \cdot 2'4] = 1'24$

Se estabiliza

Política optima.

Para A: $\pi^*(A) = p$ La que menos coste produce en Ec. Bellman.

Para B: $\pi^*(B) = q$

C no aplica.

Ejercicio 1: Solución

Modelar el MDP

- Se define mediante la tupla: $< S, A, P, C >$
  - S: Estados: $s_t \in \{A, B, C\}$
  - A: Acciones: $\{p, q\}$
  - P: Función de transición
  - $P_p\left(s_{t+1} \mid s_t\right)$:

	A	B	C
A	.2	.8	0

  - $P_q\left(s_{t+1} \mid s_t\right)$:

	A	B	C
A	0.9	0	.1
B	.1	0	.9

  - C: Coste de ejecutar cada acción
    c(p)=1
    c(q)=1

Ejercicio 1: Solución

Ecuaciones de Bellman

- Actualización de V(A):

$$V_{i+1}(A) = \text{mín}\,[c(p) + P_p(A \mid A)\, V_i(A) + P_p(B \mid A)\, V_i(B)$$
$$c(q) + P_q(C \mid A)\, V_i(C) + P_q(A \mid A)\, V_i(A)]$$

- Actualización de V(B):

$$V_{i+1}(B) = c(q) + P_q(A \mid B)\, V_i(A) + P_q(C \mid B)\, V_i(C)$$

Ejercicio 1: Solución

Calcular el valor esperado $V(s)$ para cada estado

- Iteración 0:

$$V_0(A) = 0$$

$$V_0(B) = 0$$

$$V_i(C) = 0$$

- Actualiza:

$$V_{i+1}(A) = \text{mín}\,[1 + 0.2 \times V_i(A) + 0.8 \times V_i(B) \quad 1 + 0.1 \times V_i(C) + 0.9 \times V_i(A)]$$

$$\begin{aligned} V_{i+1}(B) &= 1 + 0.1 \times V_i(A) + 0.1 \times V_i(C) \\ &= 1 + 0.1 \times V_i(A) \end{aligned}$$

- Después de 6 iteraciones los valores no se modifican:

i	0	1	2	3	4	...	N
$V_i(A)$	0	1	1.9	2.26	2.4	...	2.4
$V_i(B)$	0	1	1.1	1.19	1.2	...	1.2

Ejercicio 1: Solución

Calcular la política óptima

- Hemos obtenido los valores:
  - $V(A) = 2.4$
  - $V(B) = 1.2$
  - $V(C) = 0$
- ¿Cuál es entonces la política óptima π^*?
  - $\pi^*(A)$,
    - Para **p** tenemos:
      $c(p) + P_p(A|A)V(A) + P_p(B|A)V(B)$
      $1 + (0.2)(2.4) + (0.8)(1.2) = 2.44$
    - Para **q** tenemos:
      $c(q) + P_q(C|A)V(C) + P_q(A|A)V(A)$
      $1 + (0.1)(0) + (0.9)(2.4) = 3.16$

    Por lo tanto, $\pi^*(A)$ =p; es decir, conviene intentar ir primero a B
  - $\pi^*(B)$ =q, solo se puede aplicar una acción.
  - $\pi^*(C)$ no se define. El estado meta es un estado absorbente.

Ejercicio 2: Tratamiento de enfermedad

Para el tratamiento de un cierto tipo de enfermedad se pueden ejecutar tres acciones: cirugía, suministrar un fármaco llamado Q o un fármaco llamado R.

- Cuando la enfermedad está en su estado inicial, se pueden aplicar los tres tratamientos:
  - Si se suministra el fármaco Q, la probabilidad de curación es .3, y no tiene efecto en el resto de los casos.
  - Si se suministra el fármaco R, la probabilidad de curación es .3, con probabilidad .6 el caso se agrava y con .1 no tendrá efecto.
  - Si se decide por la cirugía (S), la probabilidad de curación es .5. Con probabilidad .4 la enfermedad se mantendrá y con probabilidad .1 se agrava.
- Para tratar un paciente en estado agravado se puede utilizar fármacos (de cualquier tipo). El fármaco R produce la curación con probabilidad .3, y el tipo Q con probabilidad .6. De lo contrario, la enfermedad sigue en el mismo estado.

El coste del fármaco R es 6, el del Q es 10 y el de la cirugía es 100. Modele el MDP y calcule la política óptima para cada estado.

Ejercicio 2)

Modelo MDP

$< S, A, P, C >$

$S = \{$ Enfermo E, Curado B (Meta), Grave $G \}$

$A = \{$ Cirugía C, R, $Q \}$

$P = \{$

$P_c(X/Y)$	E	B	G
E	0'4	0'5	0'1
G	0	0	1

$P_R(X/Y)$	E	B	G
E	0'1	0'3	0'6
G	0	0'3	0'7

$P_Q(X/Y)$	E	B	G
E	0'7	0.3	0
G	0	0.6	0'4

$\}$

$C = \{ C(R)=6,\ C(Q)=10,\ C(C)=100 \}$

Política optima

1º Hallar ec. de Bellman para ver mejores opciones.

$$V_{i+1}(E) = \min\Big[C(C) + P_c(E/E)V_i(E) + P_c(B/E)V_i(B) + P_c(G/E)V_i(G),$$
$$C(R) + P_R(E/E)V_i(E) + P_R(B/E)V_i(B) + P_R(G/E)V_i(G),$$
$$C(Q) + P_Q(E/E)V_i(E) + P_Q(B/E)V_i(B)\Big]$$

$$V_{i+1}(G) = \min\Big[C(C) + P_c(G/G)\cdot V_i(G),$$
$$C(R) + P_R(B/G)V_i(B) + P_R(G/G)V_i(G),$$
$$C(Q) + P_Q(B/G)V_i(B) + P_Q(G/G)V_i(G)\Big]$$

2º Encontrar cuando se estabiliza.

En un excel hecho es $V_{20}(G) = 1777$

$V_{20}(E) = 16'66$

3º Coge el R para Enfermo y Q para Grave.

Modelar el MDP

- Se define mediante la tupla: $< S, A, P, C >$
  - S: Estados: $s_t \in \{e, a, s\}$ (Enfermo, Complicación, Sano). Sano es el estado meta.
  - A: Acciones: $\{q, r, s\}$. q: fármaco Q, r: fármaco R, s cirugía.
  - C: Coste de ejecutar cada acción:
    c(r)=6, c(q)=10, c(s)=100
  - P: Función de transición:
    $P_q\,(s_{t+1} \mid s_t)$:

	e	c	s
e	0.7	0	0.3
c	0	0.4	0.6

  - $P_r\,(s_{t+1} \mid s_t)$

	e	c	s
e	0.1	0.6	0.3
c	0	0.7	0.3

  - $P_s\,(s_{t+1} \mid s_t)$

	e	c	s
e	0.4	0.1	0.5

Ecuaciones de Bellman:

- Actualizar el estado V(c):

$$V_{i+1}(c) = \text{mín}\left[s(q) + P_q(c \mid c)\, V_i(c) + P_q(s \mid m)\, V_i(s)\ ,\right.$$
$$\left. c(r) + P_r(c \mid c)\, V_i(c) + P_r(s \mid c)\, V_i(s)\right]$$

- Actualizar el estado V(e):

$$V_{i+1}(e) = \text{mín}\left[s(q) + P_q(e \mid e)\, V_i(e) + P_q(s \mid e)\, V_i(s)\ ,\right.$$
$$s(r) + P_r(e \mid e)\, V(e) + P_r(c \mid e)\, V(c) + P_r(s \mid e)\, V(s)\ ,$$
$$\left. s(s) + P_s(e \mid e)\, V(e) + P_s(c \mid e)\, V(c) + P_s(s \mid e)\, V(s)\right]$$

Tratamiento de enfermedad, Iteración de valores

Iteración 0: $V_0(e) = 0$ $V_0(c) = 0$

Tratamiento de enfermedad, Iteración de valores

Iteración 0: $V_0(e) = 0 \quad V_0(c) = 0$

$$
\begin{aligned}
V_1(e) = \text{mín}(& \\
10 + P_q(e \mid e) \times V_0(e) + P_q(c \mid e) \times V_0(c) + P_q(s \mid e) \times V_0(s) & = 10 + 0.7 \times V_0(e) + 0 \times V_0(c) + 0.3 \times V_0(s) = 10, \\
6 + P_r(e \mid e) \times V_0(e) + P_r(c \mid e) \times V_0(c) + P_r(s \mid e) \times V_0(s) & = 6 + 0.1 \times V_0(e) + 0.6 \times V_0(c) + 0.3 \times V_0(s) = 6, \\
100 + P_s(e \mid e) \times V_0(e) + P_s(c \mid e) \times V_0(c) + P_s(s \mid e) \times V_0(s) & = 100 + 0.4 \times V_0(e) + 0.1 \times V_0(c) + 0.5 \times V_0(s) = 100)
\end{aligned}
$$

Tratamiento de enfermedad, Iteración de valores

Iteración 0: $V_0(e) = 0 \quad V_0(c) = 0$
Iteración 1: $V_1(e) = 6$

Tratamiento de enfermedad, Iteración de valores

Iteración 0: $V_0(e) = 0 \quad V_0(c) = 0$
Iteración 1: $V_1(e) = 6$

$$V_1(c) = \text{mín}\left(\begin{array}{ll} 10 + P_q(e \mid c) \times V_0(e) + P_q(c \mid c) \times V_0(c) + P_q(s \mid c) \times V_0(s) & = 10 + 0 \times V_0(e) + 0.4 \times V_0(c) + 0.6 \times V_0(s) = 10, \\ 6 + P_r(e \mid c) \times V_0(e) + P_r(c \mid c) \times V_0(c) + P_r(s \mid c) \times V_0(s) & = 6 + 0 \times V_0(e) + 0.7 \times V_0(c) + 0.3 \times V_0(s) = 6) \end{array}\right.$$

Tratamiento de enfermedad, Iteración de valores

Iteración 0: $V_0(e) = 0 \quad V_0(c) = 0$
Iteración 1: $V_1(e) = 6 \quad V_1(c) = 6$

Tratamiento de enfermedad, Iteración de valores

Iteración 0: $V_0(e) = 0 \quad V_0(c) = 0$
Iteración 1: $V_1(e) = 6 \quad V_1(c) = 6$

$$
\begin{aligned}
V_2(e) = \text{mín}(& \\
10 + P_q(e \mid e) \times V_1(e) + P_q(c \mid e) \times V_1(c) + P_q(s \mid e) \times V_1(s) = \cdots & \quad = 10 + 0.7 \times 6 + 0.3 \times 0 = 14.2, \\
6 + P_r(e \mid e) \times V_1(e) + P_r(c \mid e) \times V_1(c) + P_r(s \mid e) \times V_1(s) = \cdots & \quad = 6 + 0.1 \times 6 + 0.6 \times 6 + 0.3 \times 0 = 10.2, \\
100 + P_s(e \mid e) \times V_1(e) + P_s(c \mid e) \times V_1(c) + P_s(s \mid e) \times V_1(s) = \cdots & \quad = 100 + 0.4 \times 6 + 0.1 \times 6 + 0.5 \times 0 = 103)
\end{aligned}
$$

Tratamiento de enfermedad, Iteración de valores

Iteración 0:	$V_0(e) = 0$	$V_0(c) = 0$
Iteración 1:	$V_1(e) = 6$	$V_1(c) = 6$
Iteración 2:	$V_2(e) = 10.2$	

Tratamiento de enfermedad, Iteración de valores

Iteración 0:	$V_0(e) = 0$	$V_0(c) = 0$
Iteración 1:	$V_1(e) = 6$	$V_1(c) = 6$
Iteración 2:	$V_2(e) = 10.2$	

$$V_2(c) = \text{mín}\Big(\begin{array}{ll} 10 + P_q(e \mid c) \times V_1(e) + P_q(c \mid c) \times V_1(c) + P_q(s \mid c) \times V_1(s) = \cdots & 10 + 0 \times 6 + 0.4 \times 6 + 0.6 \times 0 = 12.4, \\ 6 + P_r(e \mid c) \times V_1(e) + P_r(c \mid c) \times V_1(c) + P_r(s \mid c) \times V_1(s) = \cdots & 6 + 0 \times 6 + 0.7 \times 6 + 0.3 \times 0 = 10.2) \end{array}$$

Tratamiento de enfermedad, Iteración de valores

Iteración 0:	$V_0(e) = 0$	$V_0(c) = 0$
Iteración 1:	$V_1(e) = 6$	$V_1(c) = 6$
Iteración 2:	$V_2(e) = 10.2$	$V_2(c) = 10.2$

Tratamiento de enfermedad, Iteración de valores

Iteración 0: $V_0(e) = 0$ $V_0(c) = 0$
Iteración 1: $V_1(e) = 6$ $V_1(c) = 6$
Iteración 2: $V_2(e) = 10.2$ $V_2(c) = 10.2$

i	0	1	2	3	4	5	6	7	8	9	10
$V_i\,(e)$	0	6	10.2	13.1	15.1	16.5	17.2	17.5	17.6	17.7	17.7
$V_i\,(c)$	0	6	10.2	13.1	15.1	16	16.4	16.5	16.6	16.6	16.6

Calcular la política óptima

- Hemos obtenido los valores:
  - $V(e) = 17.7$
  - $V(c) = 16.6$
  - $V(s) = 0$
- ¿Cuál es entonces la política óptima π^*?
  - $\pi^*(e)$,
    - Para **q** tenemos: $c(q) + P_q(s|e)V(s) + P_q(e|e)V(e) = 10 + (0.3)(0) + (0.7)(17.7) = 22.39$
    - Para **r** tenemos: $c(r) + P_r(e|e)V(e) + P_r(c|e)V(c) + P_r(s|e)V(s) = 6 + (0.1)(17.7) + (0.6)(16.6) + (0.3)(0) = 17.73$
    - Para **s** tenemos: $c(s) + P_s(e|e)V(e) + P_s(c|e)V(c) + P_s(s|e)V(s) = 100 + (0.4)(17.7) + (0.1)(16.6) + (0.5)(0) = 108.74$

    Por lo tanto, $\pi^*(e)$ =r
  - $\pi^*(c)$,
    - Para **q** tenemos: $c(q) + P_q(s|c)V(s) + P_q(c|c)V(c) = 10 + (0.6)(0) + (0.4)(16.6) = 16.64$
    - Para **r** tenemos: $c(r) + P_r(s|c)V(s) + P_r(c|c)V(c) = 6 + (0.3)(0) + (0.7)(16.6) = 17.62$

    Por lo tanto, $\pi^*(c)$ =q
  - $\pi^*(s)$ no se define. El estado meta es un estado absorbente.

Ejercicio 3: Rover marciano

El sistema de planificación de un robot autónomo marciano optimiza sus acciones de forma que su consumo de energía (medido en unidades u.e.) sea mínimo. Tras realizar una misión nocturna es preciso llegar hasta un punto elevado de recarga, situado al lado contrario del cráter en cuyo borde se encuentra.
Para ello puede hacer tres cosas: deslizarse cuesta abajo, y luego subir hacia su objetivo, rodearlo por la derecha o rodearlo por la izquierda.

Ejercicio 3: Rover marciano

El sistema de planificación de un robot autónomo marciano optimiza sus acciones de forma que su consumo de energía (medido en unidades u.e.) sea mínimo. Tras realizar una misión nocturna es preciso llegar hasta un punto elevado de recarga, situado al lado contrario del cráter en cuyo borde se encuentra.
Para ello puede hacer tres cosas: deslizarse cuesta abajo, y luego subir hacia su objetivo, rodearlo por la derecha o rodearlo por la izquierda.

- Deslizarse cuesta abajo consume 2 u.e. Deslizarse lleva con certeza al fondo del cráter, pero la ascensión que debe realizar luego no siempre tiene éxito: una de cada cinco veces no se consigue y se cae de nuevo al fondo. Cada intento de ascensión supone un gasto de 3 u.e.
- Rodear por la derecha lleva con total certeza al objetivo, pero es larga (consume 7 u.e.).
- Rodear por la izquierda es un camino más corto (consume 4 u.e.), pero el terreno es traicionero y se cae al interior del cráter una de cada cuatro veces.

1. Modelar el problema con un Proceso de Decisión de Markov, especificando los estados y las probabilidades en forma de tabla.
2. Calcular los valores para cada estado usando Iteración de Valores (bastan 4 iteraciones)
3. Calcular la política óptima para el punto de partida. ¿Cuál es el consumo esperado de energía para el recorrido?

Rover marciano, representación

- Hay tres **estados**: Inicio (I), Fondo (F) y Recarga (R).

Rover marciano, representación

- Hay tres **estados**: Inicio (I), Fondo (F) y Recarga (R).
- Hay cuatro acciones (operadores):
  - **Izquierda**(i ,C(i)=4): Aplicable en Inicio . Resultado:

S	$P_i(S' = I\|S)$	$P_i(S' = F\|S)$	$P_i(S' = R\|S)$
Inicio	0.0	0.25	0.75

Rover marciano, representación

- Hay tres **estados**: Inicio (I), Fondo (F) y Recarga (R).
- Hay cuatro acciones (operadores):
  - **Izquierda**(i ,C(i)=4): Aplicable en Inicio . Resultado:

S	$P_i(S' = I\|S)$	$P_i(S' = F\|S)$	$P_i(S' = R\|S)$
Inicio	0.0	0.25	0.75

  - **Derecha**(d ,C(d)=7): Aplicable en Inicio . Resultado: $S' = R$

Rover marciano, representación

- Hay tres **estados**: Inicio (I), Fondo (F) y Recarga (R).
- Hay cuatro acciones (operadores):
  - **Izquierda**(i ,C(i)=4): Aplicable en Inicio . Resultado:

S	$P_i(S' = I\|S)$	$P_i(S' = F\|S)$	$P_i(S' = R\|S)$
Inicio	0.0	0.25	0.75

  - **Derecha**(d ,C(d)=7): Aplicable en Inicio . Resultado: $S' = R$
  - **Bajar**(b ,C(b)=2): Aplicable en Inicio . Resultado: $S' = F$

Rover marciano, representación

- Hay tres **estados**: Inicio (I), Fondo (F) y Recarga (R).
- Hay cuatro acciones (operadores):
  - **Izquierda**(i ,C(i)=4): Aplicable en Inicio . Resultado:

S	$P_i(S' = I\|S)$	$P_i(S' = F\|S)$	$P_i(S' = R\|S)$
Inicio	0.0	0.25	0.75

  - **Derecha**(d ,C(d)=7): Aplicable en Inicio . Resultado: $S' = R$
  - **Bajar**(b ,C(b)=2): Aplicable en Inicio . Resultado: $S' = F$
  - **Subir**(s ,C(s)=3): Aplicable en Fondo .

S	$P_s(S' = I\|S)$	$P_s(S' = F\|S)$	$P_s(S' = R\|S)$
F	0.0	0.2	0.8

Rover marciano, ecuaciones de Bellman

- Para el estado *Fondo*, como hay una sola acción posible, la ecuación de Bellman, teniendo en cuenta que $V_i(\mathit{Recarga}) = 0$ quedaría:

$$\begin{aligned} V_{i+1}(F) &= C(s) + P_s(S' = I|S = F) \times V_s(I) + P_s(S' = F|S = F) \times V_s(F) + P_s(S' = R|S = F) \times V_i(R) \\ &= 3 + 0 \times V_i(I) + 0.2 \times V_i(F) + 0.8 \times V_i(R) \\ &= 3 + 0.2 \times V_i(F) \end{aligned}$$

Rover marciano, ecuaciones de Bellman

- Para el estado *Fondo*, como hay una sola acción posible, la ecuación de Bellman, teniendo en cuenta que $V_i(\mathit{Recarga}) = 0$ quedaría:

$$\begin{aligned} V_{i+1}(F) &= C(s) + P_s(S' = I|S = F) \times V_s(I) + P_s(S' = F|S = F) \times V_s(F) + P_s(S' = R|S = F) \times V_i(R) \\ &= 3 + 0 \times V_i(I) + 0.2 \times V_i(F) + 0.8 \times V_i(R) \\ &= 3 + 0.2 \times V_i(F) \end{aligned}$$

- Para el estado *Inicio*, como hay tres acciones posibles, la ecuación de Bellman es:

$$\begin{aligned} V_{i+1}(I) = \text{mín}(&C(i) + P_i(S' = I|S = I) \times V_i(I) + P_i(S' = F|S = I) \times V_i(F) + P_i(S' = R|S = I) \times V_i(R), \\ &C(d) + P_d(S' = I|S = I) \times V_i(I) + P_d(S' = F|S = I) \times V_i(F) + P_d(S' = R|S = I) \times V_i(R), \\ &C(b) + P_b(S' = I|S = I) \times V_i(I) + P_b(S' = F|S = I) \times V_i(F) + P_b(S' = R|S = I) \times V_i(R)) \\ V_{i+1}(I) = \text{mín}(&4 + 0 \times V_i(I) + 0.25 \times V_i(F) + 0.75 \times V_i(R), \\ &7, \\ &2 + 0 \times V_i(I) + 1.0 \times V_i(F) + 0 \times V_i(R)) \\ V_{i+1}(I) = \text{mín}(&4 + 0.25 \times V_i(F), 7, 2 + V_i(F)) \end{aligned}$$

Rover marciano, iteración de valores

- Aplicando el método de iteración de valores obtenemos $V(I) = 4.94$ y $V(F) = 3.75$.

$$V_{i+1}(I) = \text{mín}(4 + 0.25 \times V_i(F), 7, 2 + V_i(F))$$
$$V_{i+1}(F) = 3 + 0.2 \times V_i(F)$$

	0	1	2	3	4	5
V(I)	0	$\text{mín}(3, 7, 2) = 2$	$\text{mín}(4.75, 7, 5) = 4.75$	$\text{mín}(4.90, 7, 5.60) = 4.90$	4.93	4.94
V(F)	0	3	3.60	3.72	3.74	3.75

Rover marciano, iteración de valores

- Aplicando el método de iteración de valores obtenemos $V(I) = 4.94$ y $V(F) = 3.75$.

$$V_{i+1}(I) = \text{mín}(4 + 0.25 \times V_i(F), 7, 2 + V_i(F))$$
$$V_{i+1}(F) = 3 + 0.2 \times V_i(F)$$

	0	1	2	3	4	5
V(I)	0	$\text{mín}(3, 7, 2) = 2$	$\text{mín}(4.75, 7, 5) = 4.75$	$\text{mín}(4.90, 7, 5.60) = 4.90$	4.93	4.94
V(F)	0	3	3.60	3.72	3.74	3.75

- La política óptima en *Inicio* será tomar la acción que minimiza el coste esperado:

$$\pi^*(I) = argmin_{\{i,d,b\}}(C(i) + P_i(S' = I|S = I) \times V_i(I) + P_i(S' = F|S = I) \times V_i(F) + P_i(S' = R|S = I) \times V_i(R),$$
$$C(d) + P_d(S' = I|S = I) \times V_i(I) + P_d(S' = F|S = I) \times V_i(F) + P_d(S' = R|S = I) \times V_i(R),$$
$$C(b) + P_b(S' = I|S = I) \times V_i(I) + P_b(S' = F|S = I) \times V_i(F) + P_b(S' = R|S = I) \times V_i(R))$$

  - Para la acción i, el coste es: $4 + 0.25V(F) = 4 + 0.253.75 = 4.94$
  - Para la acción d, el coste es: 7
  - Para la acción b, el coste es: $2 + V(F) = 2 + 3.75 = 5.75$

- Luego $\pi^*(I) = i$ (Izquierda) y se gastará en media 4.94 u.e.

Ejercicio 4: Seguridad informática

El sistema inteligente de gestión de seguridad informática ha detectado un virus. En este caso debe decidir qué hacer, con la intención de identificar el virus concreto y eliminarlo en el menor tiempo posible.

Una posibilidad es ejecutar un antivirus, operación que tarda 10 min en ejecutarse. Cada pasada del antivirus sobre un virus no identificado tiene una probabilidad del 20 % de eliminarlo (E) y un 30 % de identificar el tipo de virus (I) pero no eliminarlo. Si se ejecuta cuando el virus está identificado, sigue teniendo el 20 % de eliminarlo. Otra posibilidad es llamar al informático, que tarda 25 min en realizar su tarea. En este caso, si el virus está identificado, lo elimina con un 70 % de probabilidad. Si no lo está, lo elimina sólo con un 10 % , y lo identifica con un 70 %.

1. Representa el problema con un MDP, especificando claramente estados, transiciones, costes, y a qué probabilidades corresponden cada uno de los datos anteriores.
2. Escribir las ecuaciones de Bellman para cada estado. Primero hacerlo dejándolas indicadas, y luego sustituye para que queden ecuaciones numéricas más sencillas.
3. Realizar dos iteraciones del algoritmo de Iteración de Valores (además de la inicialización de los valores a cero).
4. Al cabo de un número suficiente de iteraciones, tenemos que el valor para el estado inicial (D) es 41 y para el estado Identificado (I) es 35. ¿Cuánto tiempo se estima que se tardará en resolver la incidencia? Determine cuál es la política óptima en cada uno de los estados.

Seguridad informática, representación

- Estados: D (detectado, pero no identificado), I (identificado), E (eliminado)
- Acciones: AV (pasar antivirus, coste 10) y INF (llamar informático, coste 25)
- Transiciones:

Acción : AV

	$P_{AV}(S_{t+1}\|S_t)$		
S_t	$S_{t+1}=D$	$S_{t+1}=I$	$S_{t+1}=E$
$S=D$	0.5	0.3	0.2
$S=I$		0.8	0.2

Acción : INF

	$P_{INF}(S_{t+1}\|S_t)$		
S_t	$S_{t+1}=D$	$S_{t+1}=I$	$S_{t+1}=E$
$S=D$	0.2	0.7	0.1
$S=I$		0.3	0.7

Para el estado D, su valor V(D) se calcula:

$$
\begin{aligned}
V_{t+1}(D) = \text{mín}\{&C(AV) + P_{AV}(S_{t+1} = D|S_t = D) \cdot V_t(D) + P_{AV}(S_{t+1} = I|S_t = D) \cdot V_t(I) + P_{AV}(S_{t+1} = E|S_t = D) \cdot V_t(E),\\
&C(INF) + P_{INF}(S_{t+1} = D|S_t = D) \cdot V_t(D) + P_{INF}(S_{t+1} = I|S_t = D) \cdot V_t(I) + P_{INF}(S_{t+1} = E|S_t = D) \cdot V_t(E)\}\\
V_{t+1}(D) = \text{mín}\{&10 + 0.50 \cdot V_t(D) + 0.30 \cdot V_t(I) + 0.20 \cdot V_t(E),\\
&25 + 0.20 \cdot V_t(D) + 0.70 \cdot V_t(I) + 0.10 \cdot V_t(E)\}\\
V_{t+1}(D) = \text{mín}\{&10 + 0.50 \cdot V_t(D) + 0.30 \cdot V_t(I),\\
&25 + 0.20 \cdot V_t(D) + 0.70 \cdot V_t(I)\}
\end{aligned}
$$

Para el estado I, su valor V(I) se calcula:

$$
\begin{aligned}
V_{t+1}(I) = \text{mín}\{&C(AV) + P_{AV}(S_{t+1} = D|S_t = I) \cdot V(NI) + P_{AV}(S_{t+1} = I|S_t = I) \cdot V(I) + P_{AV}(S_{t+1} = E|S_t = I) \cdot V(E),\\
&C(INF) + P_{INF}(S_{t+1} = D|S_t = I) \cdot V(NI) + P_{INF}(S_{t+1} = I|S_t = I) \cdot V(I) + P_{INF}(S_{t+1} = E|S_t = I) \cdot V(E)\}\\
V_{t+1}(I) = \text{mín}\{&10 + 0.00 \cdot V(D) + 0.80 \cdot V(I) + 0.20 \cdot V(E),\\
&25 + 0.00 \cdot V(D) + 0.30 \cdot V(I) + 0.70 \cdot V(E)\}\\
V_{t+1}(I) = \text{mín}\{&10 + 0.80 \cdot V(I),\\
&25 + 0.30 \cdot V(I)\}
\end{aligned}
$$

Seguridad informática, Iteración de valores

- En la figura vemos el resultado de aplicar las ecuaciones anteriores sucesivamente.

	Detectado (D)			*Identificado (I)*		
Iteración	**Acción: AV**	**Acción: INF**	**Min**	**Acción: AV**	**Acción: INF**	**Min**
0	0	0	0	0	0	0
1	10	25	10	10	25	10
2	18	34	18	18	28	18
3	24.4	41.2	24.4	24.4	30.4	24.4
4	29.52	46.96	29.52	29.52	32.32	29.52
5	33.62	51.57	33.62	33.62	33.86	33.62
6	36.89	55.25	36.89	36.89	35.08	35.08
7	38.97	56.94	38.97	38.07	35.53	35.53
8	40.14	57.66	40.14	38.42	35.66	35.66
9	40.77	57.99	40.77	38.53	35.7	35.7
10	41.09	58.14	41.09	38.56	35.71	35.71
11	41.26	58.22	41.26	38.57	35.71	35.71
12	41.34	58.25	41.34	38.57	35.71	35.71
13	41.39	58.27	41.39	38.57	35.71	35.71
14	41.41	58.28	41.41	38.57	35.71	35.71
15	41.42	58.28	41.42	38.57	35.71	35.71
16	41.42	58.28	41.42	38.57	35.71	35.71
17	41.43	58.28	41.43	38.57	35.71	35.71
18	41.43	58.29	41.43	38.57	35.71	35.71

DATO			**41**			**35**
TEST	41	57.7	41	38	35.5	35.5
POLÍTICA	**Acción: Antivirus**			**Acción: Informático**		

Seguridad informática, Política óptima

- El tiempo esperado para resolver la incidencia es el valor del estado inicial, es decir: $V_{t+1}(D) = 41$.
- Para la política en un estado, reemplazamos los valores que nos dan como dato en la ecuación del valor de dicho estado, y decidimos por la acción que da menor coste.
- Para el estado D:

$$\begin{aligned}
\pi^*(D) = argmin_{(AV,INF)}\{&10 + 0.50 \cdot V_t(D) + 0.30 \cdot V_t(I) && (\textbf{AV}),\\
&25 + 0.20 \cdot V_t(D) + 0.70 \cdot V_t(I) && (\textbf{INF})\}\\
\pi^*(D) = argmin_{(AV,INF)}\{&10 + 0.50 \cdot 41 + 0.30 \cdot 35 && (\textbf{AV}),\\
&25 + 0.20 \cdot 41 + 0.70 \cdot 35) && (\textbf{INF})\}\\
\pi^*(D) = argmin_{(AV,INF)}\{&41, 57.4\} && = \textbf{AV}
\end{aligned}$$

- Para el estado I:

$$\begin{aligned}
\pi^*(I) = argmin_{(AV,INF)}\{&10 + 0.80 \cdot V(I) && (\textbf{AV}),\\
&25 + 0.30 \cdot V(I) && (\textbf{INF})\}\\
\pi^*(I) = argmin_{(AV,INF)}\{&10 + 0.80 \cdot 35 && (\textbf{AV}),\\
&25 + 0.30 \cdot 35 && (\textbf{INF})\}\\
\pi^*(I) = argmin_{(AV,INF)}\{&38, 35.5\} && = \textbf{INF}
\end{aligned}$$

- El resultado es $\pi^*(D) = AV$, y $\pi^*(I) = INF$