uc3m | Universidad Carlos III de Madrid

Ejercicios Modelos de Markov

Procesos de Decisión de Markov (MDP)

Inteligencia Artificial Colmenarejo

Curso 2022-2023

Un MDP es una estrategia de Machine Learning que utiliza aprendizaje por refuerzo para premiar las acciones que te llevan por el mejor camino en un espacio de búsqueda.

Un MDP se define mediante la tupla: $< S, A, P, C >$ y el diagrama de transiciones:

- S: Estados: Lugares en donde, para un tempo i de la ejecución, se puede encontrar.
- A: Acciones: acciones que permiten cambiar de estados, o permanecer en el mismo.
- C: Coste de ejecutar cada acción.
- P: Función de transición. La probabilidad de, desde cada estado y con una acción, llegar a cada uno de los posibles estados.

El diagrama de transiciones representa de forma gráfica el MDP.

Para resolver un MDP hay que simular una ejecución hasta el infinito hasta que converja el valor esperado de cada estado. Se representa como $V_i(E)$, siendo i un instante de tiempo y E el estado calculado.

Las ecuaciones de Bellman nos servirán para encontrar dicho mejor valor posible. Según el problema, se busca maximizar o minimizar un resultado. Si tenemos recompensas al llegar a ciertos estados, buscaremos el máximo al realizar acciones, mientras que si lo que tenemos es costes por hacer acciones tendremos que minimizar.
Estas son las ecuaciones según la variante.

$$V_{i+1}(E_{orig}) = \max_{accion} [\, recom(accion) + \sum_{E_{dest}} P_{accion}(E_{dest}|E_{orig}) \cdot V_i(E_{dest})]$$

$$V_{i+1}(E_{orig}) = \min_{accion} [\, coste(accion) + \sum_{E_{dest}} P_{accion}(E_{dest}|E_{orig}) \cdot V_i(E_{dest})]$$

Calculando $V_0(E), V_1(E), V_2(E)$ … se termina alcanzando el valor donde converge. Dicho valor esperado es utilizado para encontrar la política óptima, es decir, la mejor acción probabilísticamente calculada para estado E. Se representa como: $\boldsymbol{\pi^*(E)}$

Ejercicio 1

Enunciado

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.

Preguntas:

- Modelar formalmente el Markov Decision Problem (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

Solución – Modelar el MDP

Se define mediante la tupla: $< S, A, P, C >$ y el diagrama de transiciones

S: **Estados**:

A: **Acciones**:

P: **Función de transición**

C: **Coste** de ejecutar cada acción

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.

Ejercicio 1

Solución – Modelar el MDP

Se define mediante la tupla: $< S, A, P, C >$ y el diagrama de transiciones

S: **Estados**: $S_t \in \{A, B, C\}$

A: **Acciones**: $\{p, q\}$

P: **Función de transición**

$P_p(S_{t+1}|S_t)$:

$P_q(S_{t+1}|S_t)$:

C: **Coste** de ejecutar cada acción

- $c(p) = 1$
- $c(q) = 1$

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.

uc3m

Solución – Modelar el MDP

Se define mediante la tupla: $< S, A, P, C >$ y el diagrama de transiciones

S: **Estados**: $S_t \in \{A, B, C\}$

A: **Acciones**: $\{p, q\}$

P: **Función de transición**

$P_p(S_{t+1}|S_t)$:

	A	B	C
A	0,2	0,8	0

$P_q(S_{t+1}|S_t)$:

C: **Coste** de ejecutar cada acción

- $c(p) = 1$
- $c(q) = 1$

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.

Ejercicio 1

Solución – Modelar el MDP

Se define mediante la tupla: $< S, A, P, C >$ y el diagrama de transiciones

S: **Estados**: $S_t \in \{A, B, C\}$

A: **Acciones**: $\{p, q\}$

P: **Función de transición**

$P_p(S_{t+1}|S_t)$:

	A	**B**	**C**
A	0,2	0,8	0

$P_q(S_{t+1}|S_t)$:

	A	**B**	**C**
A	0,9	0	0,1
B	0,1	0	0,9

C: **Coste** de ejecutar cada acción

- $c(p) = 1$
- $c(q) = 1$

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.

Solución – Modelar el MDP

Se define mediante la tupla: $< S, A, P, C >$ y el diagrama de transiciones

S: **Estados**: $S_t \in \{A, B, C\}$

A: **Acciones**: $\{p, q\}$

P: **Función de transición**

$P_p(S_{t+1}|S_t)$:

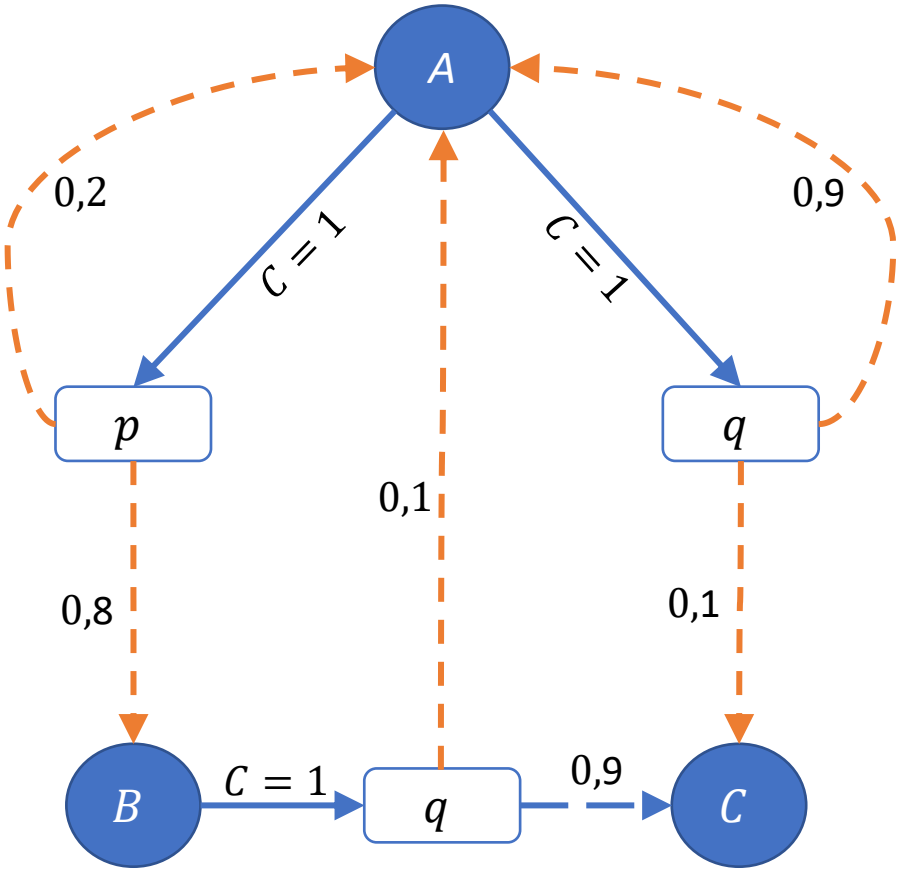
	A	B	C
A	0,2	0,8	0

$P_q(S_{t+1}|S_t)$:

	A	B	C
A	0,9	0	0,1
B	0,1	0	0,9

C: **Coste** de ejecutar cada acción

- $c(p) = 1$
- $c(q) = 1$



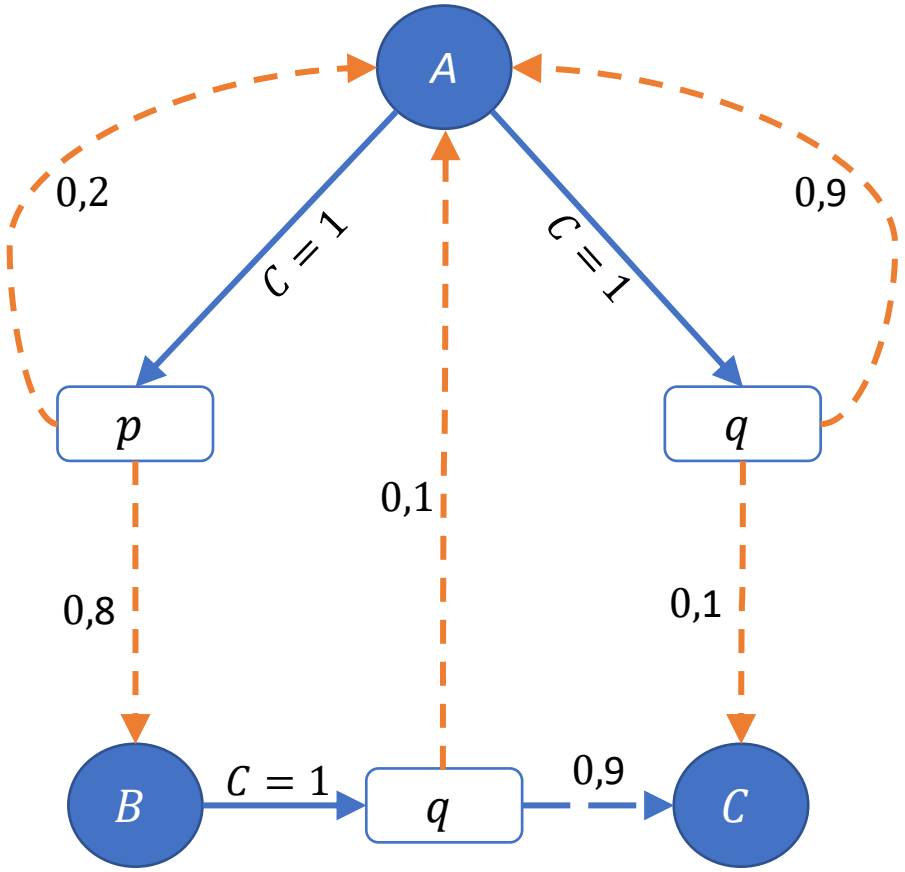
uc3m

Solución – Ecuaciones de Bellman

$$V_{i+1}(E_{orig}) = \max_{accion} [\, recom(accion) + \sum_{E_{dest}} P_{accion}(E_{dest}|E_{orig}) \cdot V_i(E_{dest})]$$
$$V_{i+1}(E_{orig}) = \min_{accion} [\, coste(accion) + \sum_{E_{dest}} P_{accion}(E_{dest}|E_{orig}) \cdot V_i(E_{dest})]$$

Actualización de V(A):

Actualización de V(B):



uc3m

Solución – Ecuaciones de Bellman

$$V_{i+1}(E_{orig}) = \max_{accion} [\, recom(accion) + \sum_{E_{dest}} P_{accion}(E_{dest}|E_{orig}) \cdot V_i(E_{dest})]$$

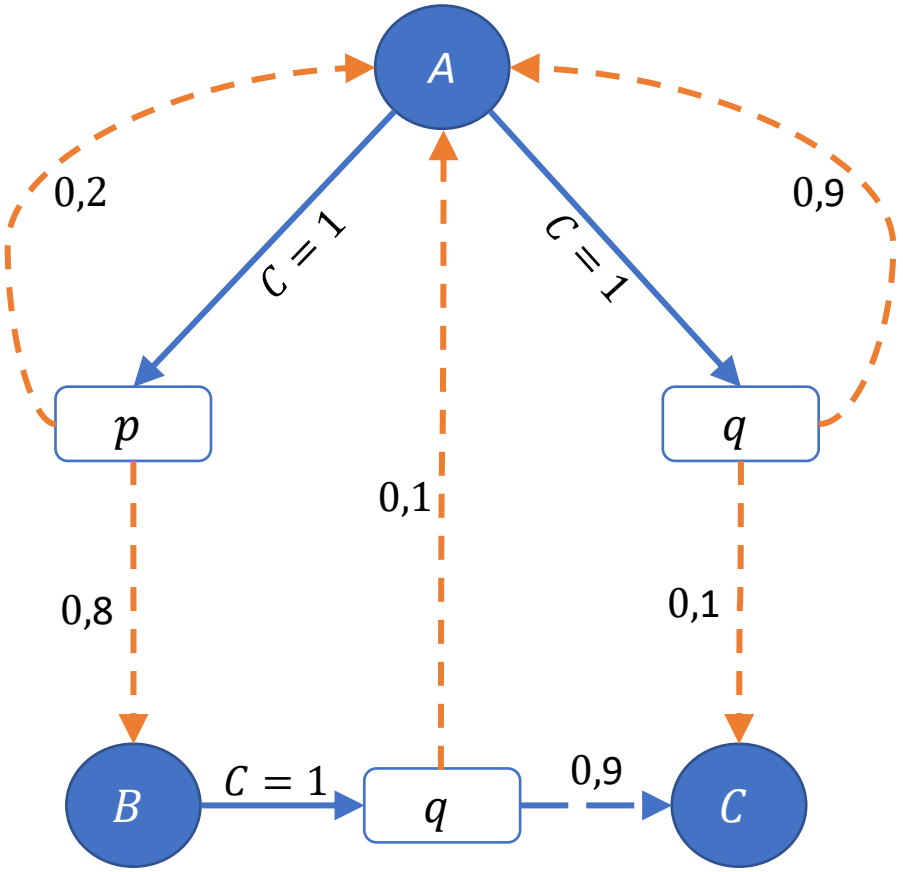
$$V_{i+1}(E_{orig}) = \min_{accion} [\, coste(accion) + \sum_{E_{dest}} P_{accion}(E_{dest}|E_{orig}) \cdot V_i(E_{dest})]$$

Actualización de V(A):

$$c(p) + P_p(A|A) \cdot V_i(A) + P_p(B|A) \cdot V_i(B)$$

$$c(q) + P_q(C|A) \cdot V_i(C) + P_q(A|A) \cdot V_i(A)$$

Actualización de V(B):



uc3m

Solución – Ecuaciones de Bellman

$$V_{i+1}(E_{orig}) = \max_{accion} [\, recom(accion) + \sum_{E_{dest}} P_{accion}(E_{dest}|E_{orig}) \cdot V_i(E_{dest})]$$
$$V_{i+1}(E_{orig}) = \min_{accion} [\, coste(accion) + \sum_{E_{dest}} P_{accion}(E_{dest}|E_{orig}) \cdot V_i(E_{dest})]$$

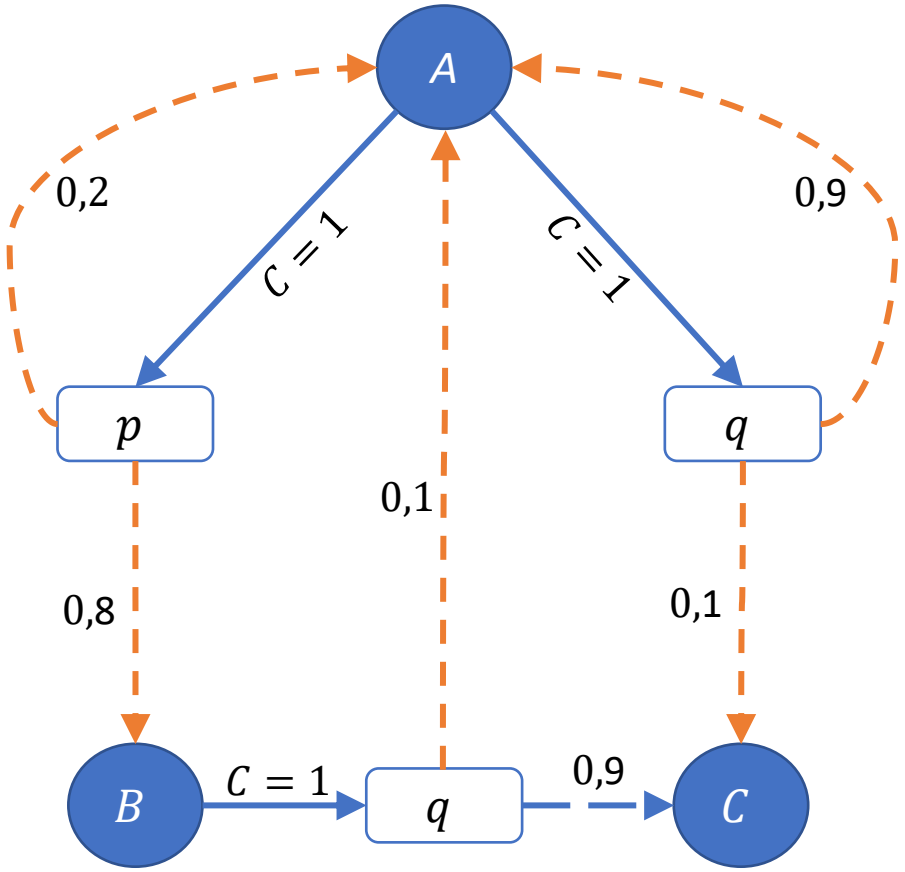
Actualización de V(A):

$$c(p) + P_p(A|A) \cdot V_i(A) + P_p(B|A) \cdot V_i(B)$$

$$c(q) + P_q(C|A) \cdot V_i(C) + P_q(A|A) \cdot V_i(A)$$

$$V_{i+1}(A) = min\,[$$
$$1 + 0{,}2 \cdot V_i(A) + 0{,}8 \cdot V_i(B)\,,$$
$$1 + 0{,}1 \cdot V_i(C) + 0{,}9 \cdot V_i(A)$$
$$]$$

Actualización de V(B):



uc3m

Solución – Ecuaciones de Bellman

$$V_{i+1}(E_{orig}) = \max_{accion} [\, recom(accion) + \sum_{E_{dest}} P_{accion}(E_{dest}|E_{orig}) \cdot V_i(E_{dest})]$$
$$V_{i+1}(E_{orig}) = \min_{accion} [\, coste(accion) + \sum_{E_{dest}} P_{accion}(E_{dest}|E_{orig}) \cdot V_i(E_{dest})]$$

Actualización de V(A):

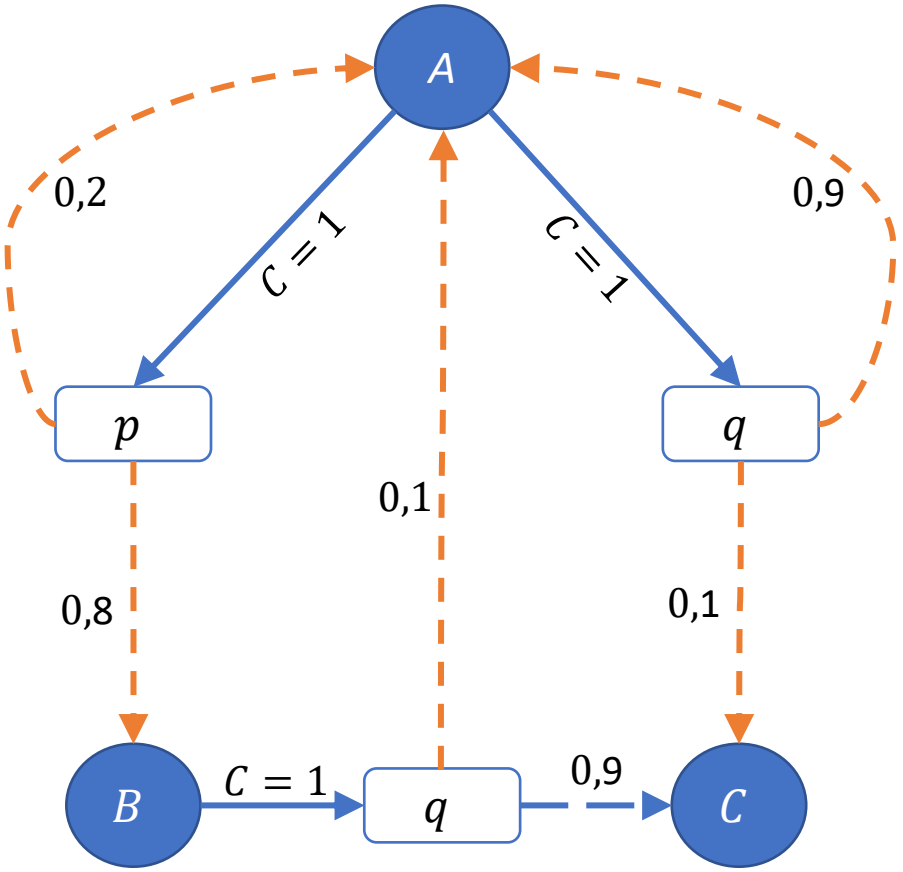
$$c(p) + P_p(A|A) \cdot V_i(A) + P_p(B|A) \cdot V_i(B)$$

$$c(q) + P_q(C|A) \cdot V_i(C) + P_q(A|A) \cdot V_i(A)$$

$$V_{i+1}(A) = min\,[$$
$$1 + 0{,}2 \cdot V_i(A) + 0{,}8 \cdot V_i(B)\,,$$
$$1 + 0{,}1 \cdot V_i(C) + 0{,}9 \cdot V_i(A)$$
$$]$$

Actualización de V(B):

$$c(q) + P_q(A|B) \cdot V_i(A) + P_q(C|B) \cdot V_i(C)$$



uc3m

Solución – Ecuaciones de Bellman

$$V_{i+1}(E_{orig}) = \max_{accion} [\, recom(accion) + \sum_{E_{dest}} P_{accion}(E_{dest}|E_{orig}) \cdot V_i(E_{dest})]$$
$$V_{i+1}(E_{orig}) = \min_{accion} [\, coste(accion) + \sum_{E_{dest}} P_{accion}(E_{dest}|E_{orig}) \cdot V_i(E_{dest})]$$

Actualización de V(A):

$$c(p) + P_p(A|A) \cdot V_i(A) + P_p(B|A) \cdot V_i(B)$$

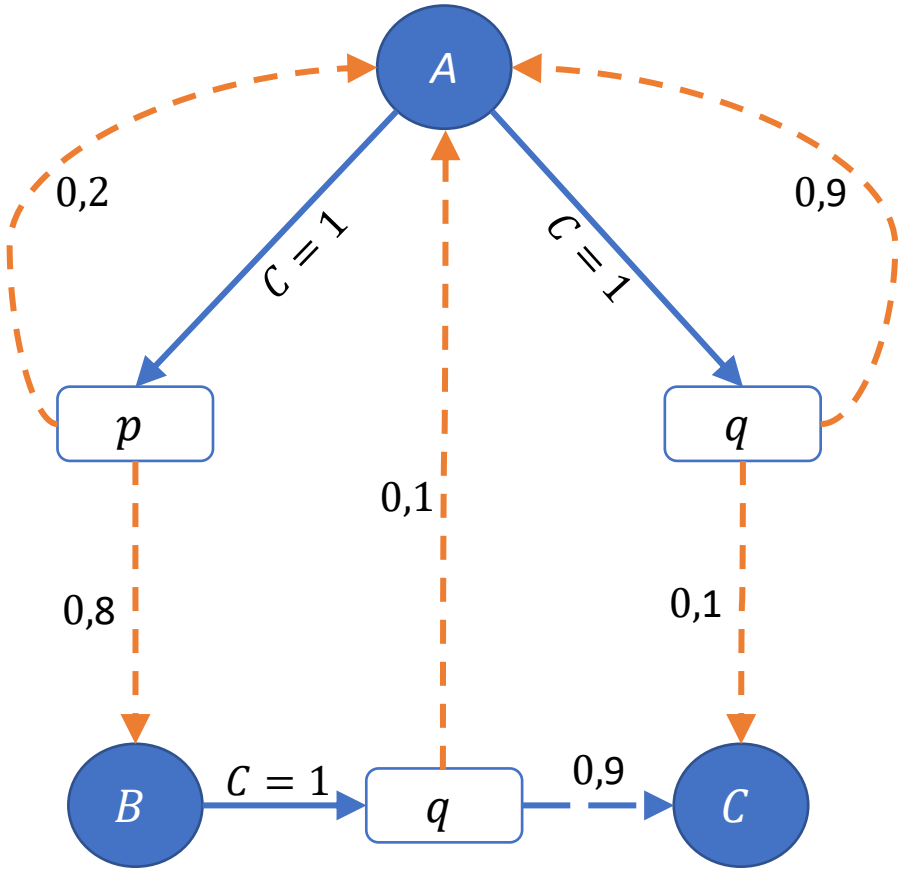
$$c(q) + P_q(C|A) \cdot V_i(C) + P_q(A|A) \cdot V_i(A)$$

$$V_{i+1}(A) = min\,[\; 1 + 0{,}2 \cdot V_i(A) + 0{,}8 \cdot V_i(B)\,,\; 1 + 0{,}1 \cdot V_i(C) + 0{,}9 \cdot V_i(A) \;]$$

Actualización de V(B):

$$c(q) + P_q(A|B) \cdot V_i(A) + P_q(C|B) \cdot V_i(C)$$

$$V_{i+1}(B) = 1 + 0{,}1 \cdot V_i(A) + 0{,}9 \cdot V_i(C)$$



uc3m

Solución – Ecuaciones de Bellman

$$V_{i+1}(E_{orig}) = \max_{accion}[\, recom(accion) + \sum_{E_{dest}} P_{accion}(E_{dest}|E_{orig}) \cdot V_i(E_{dest})]$$
$$V_{i+1}(E_{orig}) = \min_{accion}[\, coste(accion) + \sum_{E_{dest}} P_{accion}(E_{dest}|E_{orig}) \cdot V_i(E_{dest})]$$

Actualización de V(A):

$$c(p) + P_p(A|A) \cdot V_i(A) + P_p(B|A) \cdot V_i(B)$$

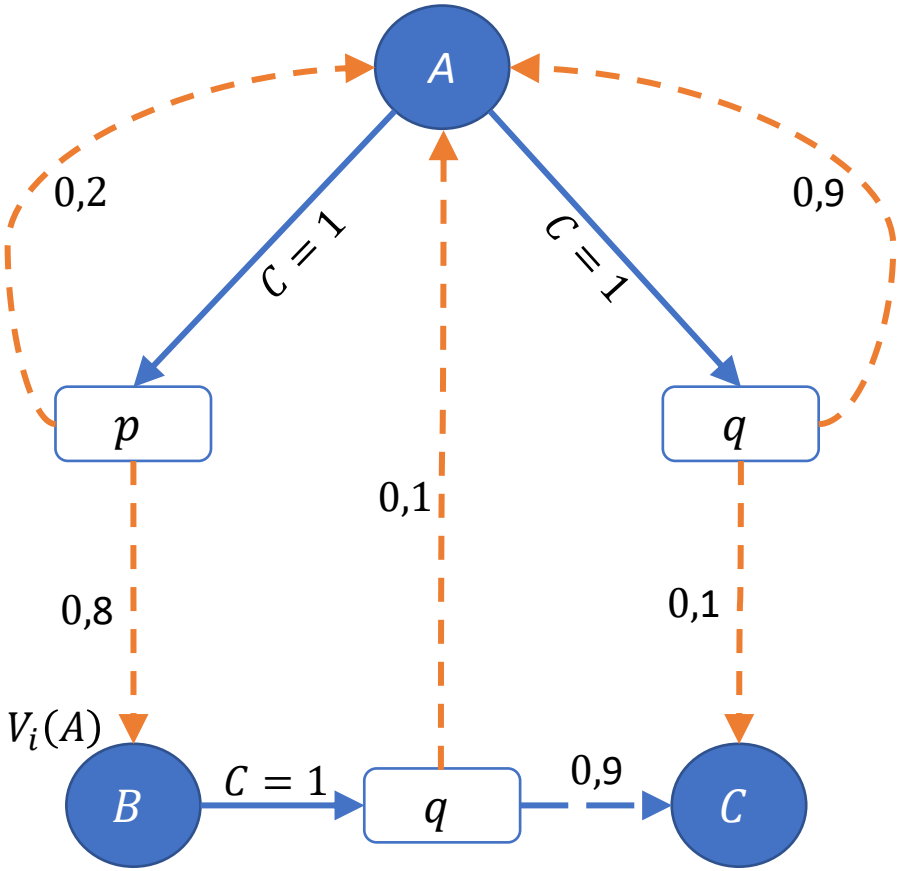
$$c(q) + P_q(C|A) \cdot V_i(C) + P_q(A|A) \cdot V_i(A)$$

$$V_{i+1}(A) = min\,[$$
$$1 + 0{,}2 \cdot V_i(A) + 0{,}8 \cdot V_i(B)\,,$$
$$1 + \cancel{0{,}1 \cdot V_i(C)} + 0{,}9 \cdot V_i(A) = 1 + 0{,}9 \cdot V_i(A)$$
$$]$$

Actualización de V(B):

$$c(q) + P_q(A|B) \cdot V_i(A) + P_q(C|B) \cdot V_i(C)$$

$$V_{i+1}(B) = 1 + 0{,}1 \cdot V_i(A) + \cancel{0{,}9 \cdot V_i(C)} = 1 + 0{,}1 \cdot V_i(A)$$



uc3m

Solución – Ecuaciones de Bellman

$$V_{i+1}(A) = min\,[\,1 + 0{,}2 \cdot V_i(A) + 0{,}8 \cdot V_i(B)\,,\,1 + 0{,}9 \cdot V_i(A)\,]$$
$$V_{i+1}(B) = 1 + 0{,}1 \cdot V_i(A)$$

Calcular el valor esperado V(s) para cada estado

- Iteración 0: $V_0(A) = 0, V_0(B) = 0, V_0(C) = 0$

$$V_0(A) = 0$$
$$V_0(B) = 0$$

$V_i(C)$ siempre será 0 porque ninguna acción sale de dicho estado y por tanto nunca es origen.
Recordemos que las ecuaciones son: $V_{i+1}(E_{orig})$=....

i	0	1	2	3	4	5	…	N
$V_i(A)$								
$V_i(B)$								

Solución – Ecuaciones de Bellman

$$V_{i+1}(A) = min\,[\,1 + 0{,}2 \cdot V_i(A) + 0{,}8 \cdot V_i(B)\,,\, 1 + 0{,}9 \cdot V_i(A)\,]$$
$$V_{i+1}(B) = 1 + 0{,}1 \cdot V_i(A)$$

Calcular el valor esperado V(s) para cada estado

- Iteración 0: $V_0(A) = 0, V_0(B) = 0, V_0(C) = 0$

$$V_1(A) = min\,[\,1 + 0{,}2 \cdot 0 + 0{,}8 \cdot 0 \quad ; \quad 1 + 0{,}9 \cdot 0] = 1$$
$$V_1(B) = 1 + 0{,}1 \cdot 0 = 1$$

i	0	1			
$V_i(A)$	0	?			
$V_i(B)$	0	?			

Ejercicio 1

Solución – Ecuaciones de Bellman

$$V_{i+1}(A) = min\,[\,1 + 0{,}2 \cdot V_i(A) + 0{,}8 \cdot V_i(B)\,,\, 1 + 0{,}9 \cdot V_i(A)\,]$$
$$V_{i+1}(B) = 1 + 0{,}1 \cdot V_i(A)$$

Calcular el valor esperado V(s) para cada estado

- Iteración 0: $V_0(A) = 0, V_0(B) = 0, V_0(C) = 0$

$$V_2(A) = min\,[\,1 + 0{,}2 \cdot 1 + 0{,}8 \cdot 1 \quad ; \quad 1 + 0{,}9 \cdot 1] = 1{,}9$$
$$V_2(B) = 1 + 0{,}1 \cdot 1 = 1{,}1$$

i	0	1	2		
$V_i(A)$	0	1	?		
$V_i(B)$	0	1	?		

Ejercicio 1

Solución – Ecuaciones de Bellman

$$V_{i+1}(A) = min\,[\,1 + 0{,}2 \cdot V_i(A) + 0{,}8 \cdot V_i(B)\,,\, 1 + 0{,}9 \cdot V_i(A)\,]$$
$$V_{i+1}(B) = 1 + 0{,}1 \cdot V_i(A)$$

Calcular el valor esperado V(s) para cada estado

- Iteración 0: $V_0(A) = 0, V_0(B) = 0, V_0(C) = 0$

$$V_3(A) = min\,[\,1 + 0{,}2 \cdot 1{,}9 + 0{,}8 \cdot 1{,}1 \quad ;\quad 1 + 0{,}9 \cdot 1{,}9] = 2{,}26$$
$$V_3(B) = 1 + 0{,}1 \cdot 1{,}9 = 1{,}19$$

i	0	1	2	3		
$V_i(A)$	0	1	1,9	?		
$V_i(B)$	0	1	1,1	?		

Solución – Ecuaciones de Bellman

$$V_{i+1}(A) = min\,[\,1 + 0{,}2 \cdot V_i(A) + 0{,}8 \cdot V_i(B)\,,\, 1 + 0{,}9 \cdot V_i(A)\,]$$
$$V_{i+1}(B) = 1 + 0{,}1 \cdot V_i(A)$$

Calcular el valor esperado V(s) para cada estado

- Iteración 0: $V_0(A) = 0, V_0(B) = 0, V_0(C) = 0$

$$V_4(A) = min\,[\,1 + 0{,}2 \cdot 2{,}26 + 0{,}8 \cdot 1{,}19 \;\;;\;\; 1 + 0{,}9 \cdot 2{,}26] = 2{,}4$$
$$V_4(B) = 1 + 0{,}1 \cdot 2{,}26 = 1{,}22$$

i	0	1	2	3	4
$V_i(A)$	0	1	1,9	2,26	?
$V_i(B)$	0	1	1,1	1,19	?

Ejercicio 1

Solución – Ecuaciones de Bellman

$$V_{i+1}(A) = min\,[\,1 + 0{,}2 \cdot V_i(A) + 0{,}8 \cdot V_i(B)\,,\,1 + 0{,}9 \cdot V_i(A)\,]$$
$$V_{i+1}(B) = 1 + 0{,}1 \cdot V_i(A)$$

Calcular el valor esperado V(s) para cada estado

- Iteración 0: $V_0(A) = 0, V_0(B) = 0, V_0(C) = 0$

$$V_{i>8}(A) = min\,[\,1 + 0{,}2 \cdot 2{,}5 + 0{,}8 \cdot 1{,}25 \quad , \quad 1 + 0{,}9 \cdot 2{,}5] = 2{,}5$$
$$V_{i>8}(B) = 1 + 0{,}1 \cdot 2{,}5 = 1{,}25$$

Después de 4 iteraciones las ecuaciones convergen y los valores no se modifican:

i	0	1	2	3	4	…	8	9
$V_i(A)$	0	1	1,9	2,26	2,4	…	2,5	2,5
$V_i(B)$	0	1	1,1	1,19	1,23	…	1,25	1,25

uc3m

Solución – Calcular la política óptima

Hemos obtenido los valores: $V(A) = 2{,}5 \quad V(B) = 1{,}25 \; V(C) = 0$

¿Cuál es entonces la política óptima?

$$\boldsymbol{\pi^*(A)} \qquad \boldsymbol{\pi^*(B)} \qquad \boldsymbol{\pi^*(C)}$$

Es decir, que acción es preferible en cada estado viendo que hacia el infinito tenemos los valores esperados: $V(A) = 2{,}5 \; V(B) = 1{,}25 \; V(C) = 0$

uc3m

Solución – Calcular la política óptima

Hemos obtenido los valores: $V(A) = 2{,}5 \quad V(B) = 1{,}25 \quad V(C) = 0$

¿Cuál es entonces la política óptima?

$$¿\,\boldsymbol{\pi}^*(\boldsymbol{A})?$$

Para **p** tenemos:

$c(p) + P_p(A|A) \cdot V(A) + P_p(B|A) \cdot V(B)$

Para **q** tenemos:

$c(q) + P_q(C|A) \cdot V(C) + P_q(A|A) \cdot V(A)$

uc3m

Solución – Calcular la política óptima

Hemos obtenido los valores: $V(A) = 2{,}5 \quad V(B) = 1{,}25 \quad V(C) = 0$

¿Cuál es entonces la política óptima?

$$¿\,\boldsymbol{\pi}^*(\boldsymbol{A})?$$

Para **p** tenemos:

$$c(p) + P_p(A|A) \cdot V(A) + P_p(B|A) \cdot V(B) = 1 + 0{,}2 \cdot 2{,}5 + 0{,}8 \cdot 1{,}25 = 2{,}5$$

Para **q** tenemos:

$$c(q) + P_q(C|A) \cdot V(C) + P_q(A|A) \cdot V(A) = 1 + 0{,}1 \cdot 0 + 0{,}9 \cdot 2{,}5 = 3{,}25$$

En el caso de A, es preferible tomar la acción p ya que se busca minimizar

$$\boldsymbol{\pi}^*(\boldsymbol{A}) = \boldsymbol{p}$$

uc3m

Solución – Calcular la política óptima

Hemos obtenido los valores: $V(A) = 2{,}5 \quad V(B) = 1{,}25 \quad V(C) = 0$

¿Cuál es entonces la política óptima?

$$¿\,\boldsymbol{\pi^*(A)}?$$

Para **p** tenemos:

$$c(p) + P_p(A|A) \cdot V(A) + P_p(B|A) \cdot V(B) = 1 + 0{,}2 \cdot 2{,}5 + 0{,}8 \cdot 1{,}25 = 2{,}5$$

Para **q** tenemos:

$$c(q) + P_q(C|A) \cdot V(C) + P_q(A|A) \cdot V(A) = 1 + 0{,}1 \cdot 0 + 0{,}9 \cdot 2{,}5 = 3{,}25$$

En el caso de A, es preferible tomar la acción p ya que se busca minimizar

$$\boldsymbol{\pi^*(A) = p}$$

En el caso de B y C:

- Solo se puede aplicar la acción q desde el estado B, por lo que:

$$\boldsymbol{\pi^*(B) = q}$$

- El estado C es un estado absorbente desde el que no se puede aplicar ninguna acción.

$$\boldsymbol{\pi^*(C)}\ \textbf{no se define}$$

Ejercicio 2

Enunciado

Para el tratamiento de un cierto tipo de tumor se pueden ejecutar tres acciones: cirugía, quimioterapia o radioterapia:

- Si se somete a quimioterapia (q), la probabilidad de curación es 0.3, regenerándose en el resto de los casos.
- Si se somete a radioterapia (r), la probabilidad de curación es 0.3, con probabilidad 0.6 se producirá metástasis y con 0.1 se regenerará.
- Si se decide extirpar (s), la probabilidad de curación es 0.5. Con probabilidad 0.4 el tumor se regenerará y con probabilidad 0.1 se producirá metástasis.
- Para tratar la metástasis se puede utilizar radioterapia o quimioterapia. La radioterapia la cura con probabilidad 0.3, y la quimioterapia con probabilidad 0.6.

El coste de la radioterapia es 6, el de la quimioterapia 10 y el de la cirugía 100.

Teniendo en cuenta que el objetivo es alcanzar la cura con el mejor coste posible:

- Modelar formalmente el MDP, no es necesario dibujar el diagrama de transiciones.
- Especificar las ecuaciones de Bellman que actualizan los valores de los estados.
- Calcular el valor esperado V(s) para cada estado.
- Calcular la política óptima.

Ejercicio 2

Solución – Modelar el MDP

Se define mediante la tupla: $< S, A, P, C >$

S: **Estados**:
A: **Acciones**:
P: **Función de transición**
C: **Coste** de ejecutar cada acción

Para el tratamiento de un cierto tipo de tumor se pueden ejecutar tres acciones: cirugía, quimioterapia o radioterapia:

- Si se somete a quimioterapia (q), la probabilidad de curación es 0.3, regenerándose en el resto de los casos.
- Si se somete a radioterapia (r), la probabilidad de curación es 0.3, con probabilidad 0.6 se producirá metástasis y con 0.1 se regenerará.
- Si se decide extirpar (s), la probabilidad de curación es 0.5. Con probabilidad 0.4 el tumor se regenerará y con probabilidad 0.1 se producirá metástasis.
- Para tratar la metástasis se puede utilizar radioterapia o quimioterapia. La radioterapia la cura con probabilidad 0.3, y la quimioterapia con probabilidad 0.6.

El coste de la radioterapia es 6, el de la quimioterapia 10 y el de la cirugía 100.

Ejercicio 2

Solución – Modelar el MDP

Se define mediante la tupla: $< S, A, P, C >$

S: **Estados**: $S_t \in \{T, M, C\}$ (tumor, metástasis, cura). Cura es el estado meta.
A: **Acciones**: $\{q, r, s\}$ q: quimioterapia, r: radioterapia, s: cirugía.
P: **Función de transición**
C: **Coste** de ejecutar cada acción

Para el tratamiento de un cierto tipo de tumor se pueden ejecutar tres acciones: cirugía, quimioterapia o radioterapia:

- Si se somete a quimioterapia (q), la probabilidad de curación es 0.3, regenerándose en el resto de los casos.
- Si se somete a radioterapia (r), la probabilidad de curación es 0.3, con probabilidad 0.6 se producirá metástasis y con 0.1 se regenerará.
- Si se decide extirpar (s), la probabilidad de curación es 0.5. Con probabilidad 0.4 el tumor se regenerará y con probabilidad 0.1 se producirá metástasis.
- Para tratar la metástasis se puede utilizar radioterapia o quimioterapia. La radioterapia la cura con probabilidad 0.3, y la quimioterapia con probabilidad 0.6.

El coste de la radioterapia es 6, el de la quimioterapia 10 y el de la cirugía 100.

Ejercicio 2

Solución – Modelar el MDP

Se define mediante la tupla: $< S, A, P, C >$

S: **Estados**: $S_t \in \{T, M, C\}$ (tumor, metástasis, cura). Cura es el estado meta.
A: **Acciones**: $\{q, r, s\}$ q: quimioterapia, r: radioterapia, s: cirugía.
P: **Función de transición**

$P_q(S_{t+1}|S_t)$:

	T	M	C
T			
M			

$P_r(S_{t+1}|S_t)$:

	T	M	C
T			
M			

$P_s(S_{t+1}|S_t)$:

	T	M	C
T			

C: **Coste** de ejecutar cada acción:
$c(r) = 6 \quad c(q) = 10 \quad c(s) = 100$

- Si se somete a quimioterapia (q), la probabilidad de curación es 0.3, regenerándose en el resto de los casos.
- Si se somete a radioterapia (r), la probabilidad de curación es 0.3, con probabilidad 0.6 se producirá metástasis y con 0.1 se regenerará.
- Si se decide extirpar (s), la probabilidad de curación es 0.5. Con probabilidad 0.4 el tumor se regenerará y con probabilidad 0.1 se producirá metástasis.
- Para tratar la metástasis se puede utilizar radioterapia o quimioterapia. La radioterapia la cura con probabilidad 0.3, y la quimioterapia con probabilidad 0.6.

Solución – Modelar el MDP

Se define mediante la tupla: $< S, A, P, C >$

S: **Estados**: $S_t \in \{T, M, C\}$ (tumor, metástasis, cura). Cura es el estado meta.
A: **Acciones**: $\{q, r, s\}$ q: quimioterapia, r: radioterapia, s: cirugía.
P: **Función de transición**

$P_q(S_{t+1}|S_t)$:

	T	M	C
T	0,7	0	0,3
M			

$P_r(S_{t+1}|S_t)$:

	T	M	C
T			
M			

$P_s(S_{t+1}|S_t)$:

	T	M	C
T			

C: **Coste** de ejecutar cada acción:
$c(r) = 6 \quad c(q) = 10 \quad c(s) = 100$

- Si se somete a quimioterapia (q), la probabilidad de curación es 0.3, regenerándose en el resto de los casos.
- Si se somete a radioterapia (r), la probabilidad de curación es 0.3, con probabilidad 0.6 se producirá metástasis y con 0.1 se regenerará.
- Si se decide extirpar (s), la probabilidad de curación es 0.5. Con probabilidad 0.4 el tumor se regenerará y con probabilidad 0.1 se producirá metástasis.
- Para tratar la metástasis se puede utilizar radioterapia o quimioterapia. La radioterapia la cura con probabilidad 0.3, y la quimioterapia con probabilidad 0.6.

Ejercicio 2

Solución – Modelar el MDP

Se define mediante la tupla: $< S, A, P, C >$

S: **Estados**: $S_t \in \{T, M, C\}$ (tumor, metástasis, cura). Cura es el estado meta.
A: **Acciones**: $\{q, r, s\}$ q: quimioterapia, r: radioterapia, s: cirugía.
P: **Función de transición**

$P_q(S_{t+1}|S_t)$:

	T	M	C
T	0,7	0	0,3
M			

$P_r(S_{t+1}|S_t)$:

	T	M	C
T	0,1	0,6	0,3
M			

$P_s(S_{t+1}|S_t)$:

	T	M	C
T			

C: **Coste** de ejecutar cada acción:
$c(r) = 6 \quad c(q) = 10 \quad c(s) = 100$

- Si se somete a quimioterapia (q), la probabilidad de curación es 0.3, regenerándose en el resto de los casos.
- Si se somete a radioterapia (r), la probabilidad de curación es 0.3, con probabilidad 0.6 se producirá metástasis y con 0.1 se regenerará.
- Si se decide extirpar (s), la probabilidad de curación es 0.5. Con probabilidad 0.4 el tumor se regenerará y con probabilidad 0.1 se producirá metástasis.
- Para tratar la metástasis se puede utilizar radioterapia o quimioterapia. La radioterapia la cura con probabilidad 0.3, y la quimioterapia con probabilidad 0.6.

Ejercicio 2

Solución – Modelar el MDP

Se define mediante la tupla: $< S, A, P, C >$

S: **Estados**: $S_t \in \{T, M, C\}$ (tumor, metástasis, cura). Cura es el estado meta.
A: **Acciones**: $\{q, r, s\}$ q: quimioterapia, r: radioterapia, s: cirugía.
P: **Función de transición**

$P_q(S_{t+1}|S_t)$:

	T	M	C
T	0,7	0	0,3
M			

$P_r(S_{t+1}|S_t)$:

	T	M	C
T	0,1	0,6	0,3
M			

$P_s(S_{t+1}|S_t)$:

	T	M	C
T	0,4	0,1	0,5

C: **Coste** de ejecutar cada acción:
$c(r) = 6 \quad c(q) = 10 \quad c(s) = 100$

- Si se somete a quimioterapia (q), la probabilidad de curación es 0.3, regenerándose en el resto de los casos.
- Si se somete a radioterapia (r), la probabilidad de curación es 0.3, con probabilidad 0.6 se producirá metástasis y con 0.1 se regenerará.
- Si se decide extirpar (s), la probabilidad de curación es 0.5. Con probabilidad 0.4 el tumor se regenerará y con probabilidad 0.1 se producirá metástasis.
- Para tratar la metástasis se puede utilizar radioterapia o quimioterapia. La radioterapia la cura con probabilidad 0.3, y la quimioterapia con probabilidad 0.6.

Ejercicio 2

Solución – Modelar el MDP

Se define mediante la tupla: $< S, A, P, C >$

S: **Estados**: $S_t \in \{T, M, C\}$ (tumor, metástasis, cura). Cura es el estado meta.
A: **Acciones**: $\{q, r, s\}$ q: quimioterapia, r: radioterapia, s: cirugía.
P: **Función de transición**

$P_q(S_{t+1}|S_t)$:

	T	M	C
T	0,7	0	0,3
M	0	0,4	0,6

$P_r(S_{t+1}|S_t)$:

	T	M	C
T	0,1	0,6	0,3
M	0	0,7	0,3

$P_s(S_{t+1}|S_t)$:

	T	M	C
T	0,4	0,1	0,5

C: **Coste** de ejecutar cada acción:
$c(r) = 6 \quad c(q) = 10 \quad c(s) = 100$

- Si se somete a quimioterapia (q), la probabilidad de curación es 0.3, regenerándose en el resto de los casos.
- Si se somete a radioterapia (r), la probabilidad de curación es 0.3, con probabilidad 0.6 se producirá metástasis y con 0.1 se regenerará.
- Si se decide extirpar (s), la probabilidad de curación es 0.5. Con probabilidad 0.4 el tumor se regenerará y con probabilidad 0.1 se producirá metástasis.
- Para tratar la metástasis se puede utilizar radioterapia o quimioterapia. La radioterapia la cura con probabilidad 0.3, y la quimioterapia con probabilidad 0.6.

uc3m

Solución – Ecuaciones de Bellman

$P_q(S_{t+1}|S_t)$:

	T	M	C
T	0,7	0	0,3
M	0	0,4	0,6

$P_r(S_{t+1}|S_t)$:

	T	M	C
T	0,1	0,6	0,3
M	0	0,7	0,3

$P_s(S_{t+1}|S_t)$:

	T	M	C
T	0,4	0,1	0,5

$c(r) = 6$
$c(q) = 10$
$c(s) = 100$

$$V_{i+1}(E_{orig}) = \min_{accion} [\, coste(accion) + \sum_{E_{dest}} P_{accion}(E_{dest}|E_{orig}) \cdot V_i(E_{dest})]$$

Actualización de V(M):

Actualización de V(T):

Actualización de V(C):

Ejercicio 2

Solución – Ecuaciones de Bellman

$P_q(S_{t+1}|S_t)$:

	T	M	C
T	0,7	0	0,3
M	0	0,4	0,6

$P_r(S_{t+1}|S_t)$:

	T	M	C
T	0,1	0,6	0,3
M	0	0,7	0,3

$P_s(S_{t+1}|S_t)$:

	T	M	C
T	0,4	0,1	0,5

$c(r) = 6$
$c(q) = 10$
$c(s) = 100$

$$V_{i+1}(E_{orig}) = \min_{accion} [\, coste(accion) + \sum_{E_{dest}} P_{accion}(E_{dest}|E_{orig}) \cdot V_i(E_{dest})]$$

Actualización de V(M):

$$V_{i+1}(M) = min\,[c(q) + P_q(M|M) \cdot V_i(M) + P_q(C|M) \cdot V_i(C)$$
$$c(r) + P_r(M|M) \cdot V_i(M) + P_r(C|M) \cdot V_i(C)\,]$$

Actualización de V(T):

$$V_{i+1}(T) = min\,[\, c(q) + P_q(T|T) \cdot V_i(T) + P_q(C|T) \cdot V_i(C),$$
$$c(r) + P_r(T|T) \cdot V_i(T) + P_r(M|T) \cdot V_i(C),$$
$$c(s) + P_s(T|T) \cdot V_i(T) + P_s(M|T) \cdot V_i(C)\,]$$

Actualización de V(C):

No tiene al ser un estado meta, es decir, $V_i(C)$=0

uc3m

Solución – Ecuaciones de Bellman

$P_q(S_{t+1}|S_t)$:

	T	M	C
T	0,7	0	0,3
M	0	0,4	0,6

$P_r(S_{t+1}|S_t)$:

	T	M	C
T	0,1	0,6	0,3
M	0	0,7	0,3

$P_s(S_{t+1}|S_t)$:

	T	M	C
T	0,4	0,1	0,5

$c(r) = 6$
$c(q) = 10$
$c(s) = 100$

$$V_{i+1}(E_{orig}) = \min_{accion} [\, coste(accion) + \sum_{E_{dest}} P_{accion}(E_{dest}|E_{orig}) \cdot V_i(E_{dest})]$$

Actualización de V(M):

$$V_{i+1}(M) = min\,[c(q) + P_q(M|M) \cdot V_i(M) + \cancel{P_q(C|M) \cdot V_i(C)}$$
$$c(r) + P_r(M|M) \cdot V_i(M) + \cancel{P_r(C|M) \cdot V_i(C)}\,]$$

Actualización de V(T):

$$V_{i+1}(T) = min\,[\, c(q) + P_q(T|T) \cdot V_i(T) + \cancel{P_q(C|T) \cdot V_i(C)},$$
$$c(r) + P_r(T|T) \cdot V_i(T) + \cancel{P_r(M|T) \cdot V_i(C)},$$
$$c(s) + P_s(T|T) \cdot V_i(T) + \cancel{P_s(M|T) \cdot V_i(C)}\,]$$

Actualización de V(C):

No tiene al ser un estado meta, es decir, $V_i(C)$=0

Ejercicio 2

Solución – Ecuaciones de Bellman

$P_q(S_{t+1}|S_t)$:

	T	M	C
T	0,7	0	0,3
M	0	0,4	0,6

$P_r(S_{t+1}|S_t)$:

	T	M	C
T	0,1	0,6	0,3
M	0	0,7	0,3

$P_s(S_{t+1}|S_t)$:

	T	M	C
T	0,4	0,1	0,5

$c(r) = 6$
$c(q) = 10$
$c(s) = 100$

$$V_{i+1}(M) = min\,[10 + 0 \cdot V_i(T) + 0{,}4 \cdot V_i(M),\; 6 + 0 \cdot V_i(T) + 0{,}7 \cdot V_i(M)]$$

$$V_{i+1}(T) = min\,[10 + 0{,}7 \cdot V_i(T) + 0 \cdot V_i(M),\; 6 + 0{,}1 \cdot V_i(T) + 0{,}6 \cdot V_i(M),\; 100 + 0{,}4 \cdot V_i(T) + 0{,}1 \cdot V_i(M)]$$

i	0	1	2	3	4	5	6	7	8	9	10
$V_i(T)$	0										
$V_i(M)$	0										

Ejercicio 2

Solución – Ecuaciones de Bellman

$P_q(S_{t+1}|S_t)$:

	T	M	C
T	0,7	0	0,3
M	0	0,4	0,6

$P_r(S_{t+1}|S_t)$:

	T	M	C
T	0,1	0,6	0,3
M	0	0,7	0,3

$P_s(S_{t+1}|S_t)$:

	T	M	C
T	0,4	0,1	0,5

$c(r) = 6$
$c(q) = 10$
$c(s) = 100$

$$V_{i+1}(M) = min\,[10 + 0 \cdot V_i(T) + 0{,}4 \cdot V_i(M),\ 6 + 0 \cdot V_i(T) + 0{,}7 \cdot V_i(M)]$$

$$V_{i+1}(T) = min\,[10 + 0{,}7 \cdot V_i(T) + 0 \cdot V_i(M),\ 6 + 0{,}1 \cdot V_i(T) + 0{,}6 \cdot V_i(M),\ 100 + 0{,}4 \cdot V_i(T) + 0{,}1 \cdot V_i(M)]$$

i	0	1	2	3	4	5	6	7	8	9	10
$V_i(T)$	0	6									
$V_i(M)$	0	6									

uc3m

Solución – Ecuaciones de Bellman

$P_q(S_{t+1}|S_t)$:

	T	M	C
T	0,7	0	0,3
M	0	0,4	0,6

$P_r(S_{t+1}|S_t)$:

	T	M	C
T	0,1	0,6	0,3
M	0	0,7	0,3

$P_s(S_{t+1}|S_t)$:

	T	M	C
T	0,4	0,1	0,5

$$c(r) = 6$$
$$c(q) = 10$$
$$c(s) = 100$$

$$V_{i+1}(M) = min\,[10 + 0 \cdot V_i(T) + 0{,}4 \cdot V_i(M),\; 6 + 0 \cdot V_i(T) + 0{,}7 \cdot V_i(M)]$$

$$V_{i+1}(T) = min\,[10 + 0{,}7 \cdot V_i(T) + 0 \cdot V_i(M),\; 6 + 0{,}1 \cdot V_i(T) + 0{,}6 \cdot V_i(M),\; 100 + 0{,}4 \cdot V_i(T) + 0{,}1 \cdot V_i(M)]$$

i	0	1	2	3	4	5	6	7	8	9	10
$V_i(T)$	0	6	10,2	13,1	15,1	16,5	17,2	17,5	17,6	17,7	17,7
$V_i(M)$	0	6	10,2	13,1	15,1	16	16,4	16,5	16,6	16,6	16,6

Ejercicio 2

Solución – Política óptima

Hemos obtenido los valores: $V(T) = 17{,}7 \quad V(m) = 16{,}6 \quad V(C) = 0$

$$V_{i+1}(M) = min\,[c(q) + P_q(M|M) \cdot V_i(M),\ c(r) + P_r(M|M) \cdot V_i(M)]$$
$$V_{i+1}(T) = min\,[\,c(q) + P_q(T|T) \cdot V_i(T),\quad c(r) + P_r(T|T) \cdot V_i(T),\ c(s) + P_s(T|T) \cdot V_i(T)]$$

$\boldsymbol{\pi^*(C)}$ no se define. El estado meta es un estado absorbente.

$\boldsymbol{\pi^*(\mathbf{M})}$

Para q tenemos:

$c(q) + P_q(M|M) \cdot V_i(M) = 10 + 0{,}4 \cdot 16{,}6 = 16{,}64$

Para r tenemos:

$c(r) + P_r(M|M) \cdot V_i(M) = 6 + 0{,}7 \cdot 16{,}6 = 17{,}62$

$\boldsymbol{\pi^*(\mathbf{T})}$

Para q tenemos:

$c(q) + P_q(T|T) \cdot V_i(T) = 10 + 0{,}7 \cdot 17{,}7 = 22{,}39$

Para r tenemos:

$c(r) + P_r(T|T) \cdot V_i(T) + P_r(M|T) \cdot V(M) =$
$= 6 + 0{,}1 \cdot 17{,}7 \ + 0{,}6 \cdot 16{,}6 = 17{,}73$

Para s tenemos:

$c(s) + P_s(T|T) \cdot V(T) + P_s(M|T) \cdot V(M) =$
$= 100 + 0{,}4 \cdot 17{,}7 \ + 0{,}1 \cdot 16{,}6 = 108{,}74$

Ejercicio 2

Solución – Política óptima

Hemos obtenido los valores: $V(T) = 17{,}7 \quad V(m) = 16{,}6 \quad V(C) = 0$

$$V_{i+1}(M) = min\,[c(q) + P_q(M|M) \cdot V_i(M),\ c(r) + P_r(M|M) \cdot V_i(M)]$$
$$V_{i+1}(T) = min\,[\,c(q) + P_q(T|T) \cdot V_i(T),\quad c(r) + P_r(T|T) \cdot V_i(T),\ c(s) + P_s(T|T) \cdot V_i(T)]$$

$\boldsymbol{\pi^*(C)}$ no se define. El estado meta es un estado absorbente.

$\boldsymbol{\pi^*(\mathbf{M})}$

Para q tenemos:

$c(q) + P_q(M|M) \cdot V_i(M) = 10 + 0{,}4 \cdot 16{,}6 = 16{,}64$

Para r tenemos:

$c(r) + P_r(M|M) \cdot V_i(M) = 6 + 0{,}7 \cdot 16{,}6 = 17{,}62$

Por lo tanto, $\boldsymbol{\pi^*(M) = \mathbf{q}}$

$\boldsymbol{\pi^*(\mathbf{T})}$

Para q tenemos:

$c(q) + P_q(T|T) \cdot V_i(T) = 10 + 0{,}7 \cdot 17{,}7 = 22{,}39$

Para r tenemos:

$c(r) + P_r(T|T) \cdot V_i(T) + P_r(M|T) \cdot V(M) =$
$= 6 + 0{,}1 \cdot 17{,}7 \ + 0{,}6 \cdot 16{,}6 = 17{,}73$

Para s tenemos:

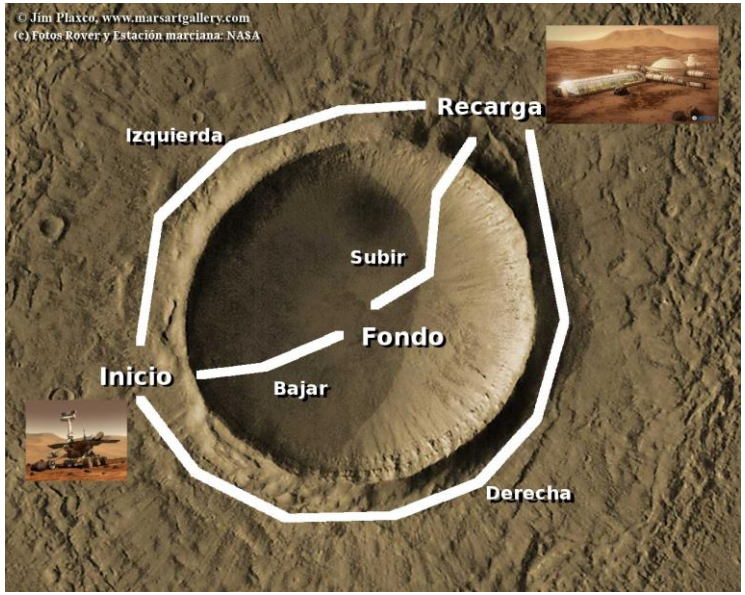
$c(s) + P_s(T|T) \cdot V(T) + P_s(M|T) \cdot V(M) =$
$= 100 + 0{,}4 \cdot 17{,}7 \ + 0{,}1 \cdot 16{,}6 = 108{,}74$

Por lo tanto, $\boldsymbol{\pi^*(T) = \mathbf{r}}$

Enunciado

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 Rover

Enunciado

- 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.

Responda a las siguientes cuestiones:

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?

Representación

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

	I	F	R
I	0	0.25	0.75

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

	I	F	R
I	0	0	1

  - Bajar(b): C(b)=2. Aplicable en Inicio . Resultado: S = F

	I	F	R
I	0	1	0

  - Subir(s): C(s)=3. Aplicable en Fondo .

	I	F	R
F	0	0.2	0.8

Ecuaciones de Bellman

Para el estado Fondo, como hay una sola acción posible, la ecuación de Bellman, teniendo en cuenta que Vi (Recarga) = 0, quedaría:

$$V_{i+1}(F) = C(s)+P_s(S=I|S=F)\cdot V_s(I)+P_s(S=F|S=F)\cdot V_s(F)+P_s(S=R|S=F)\cdot V_i(R)$$

$$= 3 + 0 \cdot V_i(I)+0.2\cdot V_i(F)+0.8\cdot V_i(R)=3+0.2\cdot V_i(F)$$

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

$$V_{i+1}(I) = \min\{$$
$$C(i)+P_i(S=I|S=I)\cdot V_i(I)+P_i(S=F|S=I)\cdot V_i(F)+P_i(S=R|S=I)\cdot V_i(R),$$
$$C(d)+P_d(S=I|S=I)\cdot V_i(I)+P_d(S=F|S=I)\cdot V_i(F)+P_d(S=R|S=I)\cdot V_i(R),$$
$$C(b)+P_b(S=I|S=I)\cdot V_i(I)+P_b(S=F|S=I)\cdot V_i(F)+P_b(S=R|S=I)\cdot V_i(R),$$
$$\}$$

$$V_{i+1}(I) = \min\{$$
$$4+0\cdot V_i(I)+0{\cdot}25\cdot V_i(F)+0{\cdot}75\cdot V_i(R),$$
$$7,$$
$$2+0\cdot V_i(I)+1\cdot V_i(F)+0\cdot V_i(R),$$
$$\}$$

$$V_{i+1}(I) = \min\{$$
$$4+0{\cdot}25\cdot V_i(F)+0{\cdot}75\cdot V_i(R),$$
$$7,$$
$$2+V_i(F),$$
$$\}$$

Ecuaciones de Bellman

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

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

	0	1	2	3	4	5
V(I)	0	mín(3, 7, 2) = 2	mín(4.75, 7, 5) = 4.75	mín(4.90, 7, 5.60) = 4.90	4.93	4.94
V(F)	0	3	3.6	3.72	3.74	3.75

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

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

- Para la acción i , el coste es: $4 + 0.25 \cdot V(F) = 4 + 0.25 + 3.75 = 4.94$
- 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 Seguridad Informática

Enunciado

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 %.

Responda a las siguientes cuestiones:

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
3. dejándolas indicadas, y luego sustituye para que queden ecuaciones numéricas más sencillas.
4. Realizar dos iteraciones del algoritmo de Iteración de Valores (además de la inicialización de los valores a cero).
5. 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 la política óptima en cada estado.

Ejercicio Seguridad Informática

Representación

Estados:

- D (detectado, pero no identificado)
- I (identificado)
- E(eliminado)

Acciones:

- AV (pasar antivirus, coste 10)
- INF (llamar informático, coste 25)

Tabla de Transiciones:

Acción : AV

$P_{AV}(S_{t+1}|S_t)$

St	St +1 = D	St +1 = I	St +1 = E
S = D	0.5	0.3	0.2
S = I		0.8	0.2

Acción : INF

$P_{INF}(S_{t+1}|S_t)$

St	St +1 = D	St +1 = I	St +1 = E
S = D	0.2	0.7	0.1
S = I		0.3	0.7

Ejercicio Seguridad Informática

Ecuaciones de Bellman

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

$$V_{t+1}(D) = mín\{$$
$$C(AV) + P_{AV}(S_{t+1} = D|S_t = D) \cdot V_t(D) + P_{AV}(S_{t+1} = I|S_t = D) \cdot V_t(I) + P_{AV}(S_{t+1} = E|S_t = D) \cdot V_t(E),$$
$$C(INF) + P_{INF}(S_{t+1} = D|S_t = D) \cdot V_t(D) + P_{INF}(S_{t+1} = I|S_t = D) \cdot V_t(I) + P_{INF}(S_{t+1} = E|S_t = D) \cdot V_t(E)$$
$$\}$$

$$V_{t+1}(D) = mín\{$$
$$10 + 0.50 \cdot V_t(D) + 0.30 \cdot V_t(I) + 0.20 \cdot V_t(E),$$
$$25 + 0.20 \cdot V_t(D) + 0.70 \cdot V_t(I) + 0.10 \cdot V_t(E)$$
$$\}$$

$$V_{t+1}(D) = mín\{$$
$$10 + 0.50 \cdot V_t(D) + 0.30 \cdot V_t(I),$$
$$25 + 0.20 \cdot V_t(D) + 0.70 \cdot V_t(I)$$
$$\}$$

Ecuaciones de Bellman

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

$$V_{t+1}(I) = mín\{$$
$$C(AV) + P_{AV}(S_{t+1} = D|S_t = I) \cdot V_t(NI) + P_{AV}(S_{t+1} = I|S_t = I) \cdot V(I) +$$
$$P_{AV}(S_{t+1} = E|S_t = I) \cdot V(E),$$
$$C(INF) + P_{INF}(S_{t+1} = D|S_t = I) \cdot V_t(NI) + P_{INF}(S_{t+1} = I|S_t = I) \cdot V(I) +$$
$$P_{INF}(S_{t+1} = E|S_t = I) \cdot V(E)$$
$$\}$$

$$V_{t+1}(I) = mín\{$$
$$10 + 0.00 \cdot V(D) + 0.80 \cdot V(I) + 0.20 \cdot V(E),$$
$$25 + 0.00 \cdot V(D) + 0.30 \cdot V(I) + 0.70 \cdot V(E),$$
$$\}$$

$$V_{t+1}(I) = mín\{$$
$$10 + 0.80 \cdot V(I),$$
$$25 + 0.30 \cdot V(I),$$
$$\}$$

Ecuaciones de Bellman

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**		

Ejercicio Seguridad Informática

Ecuaciones de Bellman

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:

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

Para el estado I:

$$\pi^*(I) = argmin(AV, INF)\{$$
$$10 + 0.80 \cdot V(I),$$
$$25 + 0.30 \cdot V(I)$$
$$\}$$
$$\pi^*(I) = argmin(AV, INF)\{$$
$$10 + 0.80 \cdot 35,$$
$$25 + 0.30 \cdot 35$$
$$\}$$
$$\pi^*(I) = argmin(AV, INF)\{$$
$$38,$$
$$35.5$$
$$\}$$

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