SCALAB

Universidad Carlos III de Madrid

Incertidumbre

En este tema

Incertidumbre

Razonamiento Probabilístico
Redes Bayesianas
Razonamiento Probabilístico en el tiempo
- Introducción
- Modelos de Markov
- Modelos de Markov Ocultos (HMMs)
- Procesos de Decisión de Markov (MDPs)
- Procesos de Decisión de Markov Parcialmente Observables (POMDPs)

Lógica Borrosa

En este tema

Incertidumbre

- Razonamiento Probabilístico
- Redes Bayesianas
- Razonamiento Probabilístico en el tiempo
  - Introducción
  - Modelos de Markov
  - Modelos de Markov Ocultos (HMMs)
  - Procesos de Decisión de Markov (MDPs)
  - Procesos de Decisión de Markov Parcialmente Observables (POMDPs)
- Lógica Borrosa

En este tema

Incertidumbre

Razonamiento Probabilístico
Redes Bayesianas
Razonamiento Probabilístico en el tiempo
- Introducción
- Modelos de Markov
- Modelos de Markov Ocultos (HMMs)
- Procesos de Decisión de Markov (MDPs)
- Procesos de Decisión de Markov Parcialmente Observables (POMDPs)

Lógica Borrosa

En este tema

Incertidumbre

Razonamiento Probabilístico
Redes Bayesianas
Razonamiento Probabilístico en el tiempo
- Introducción
- Modelos de Markov
- Modelos de Markov Ocultos (HMMs)
- Procesos de Decisión de Markov (MDPs)
- Procesos de Decisión de Markov Parcialmente Observables (POMDPs)

Lógica Borrosa

Probabilidades, Tiempo, Acciones

- **Razonamiento probabilístico en el tiempo**
  - Hasta ahora: razonamiento probabilístico en mundos estáticos
  - El mundo es dinámico ¿Cómo podemos considerar el paso del tiempo?
  - **Por ejemplo, cómo afecta el hecho de que lloviera ayer (Llueve_t) al hecho de que llueva hoy (Llueve_{t+1})**
  - Modelos de Markov (MM) (también denominados Procesos de Markov o cadenas de Markov) y Modelos de Markov ocultos (HMM)

Probabilidades, Tiempo, Acciones

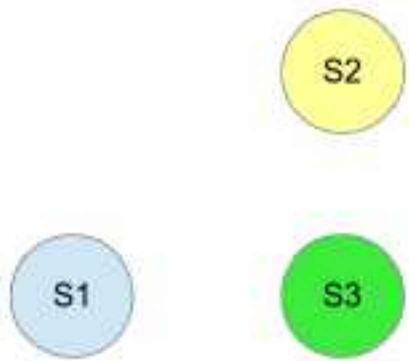
- **Razonamiento probabilístico en mundos no deterministas**
  - Los algoritmos de búsqueda proporcionan una secuencia de acciones para resolver el problema, pero
  - ¿Qué pasa si las acciones no son deterministas?
  - ¿Cómo podemos elegir las acciones que tienen más probabilidad de llevarnos al estado meta?
  - Procesos de Decisión de Markov (MDPs) y Procesos de Decisión de Markov Parcialmente Observables (POMDPs)

En este tema

Incertidumbre

- Razonamiento Probabilístico
- Redes Bayesianas
- Razonamiento Probabilístico en el tiempo
  - Introducción
  - Modelos de Markov
  - Modelos de Markov Ocultos (HMMs)
  - Procesos de Decisión de Markov (MDPs)
  - Procesos de Decisión de Markov Parcialmente Observables (POMDPs)
- Lógica Borrosa

Modelos de Markov

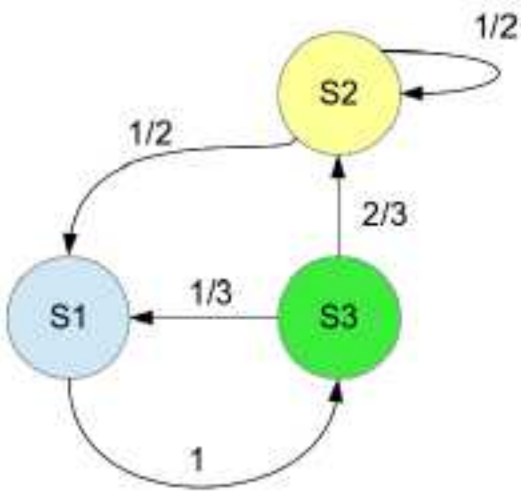


- N estados: $s_1, s_2, \ldots, s_N$
- Instantes de tiempo discretos: $t = 0, t = 1, \ldots$
- En un instante t el sistema está en un estado E_t

$$E_t \in \{s1, s2, ..., s_N\}$$

- El estado actual determina la distribución de probabilidad del estado siguiente

Modelos de Markov

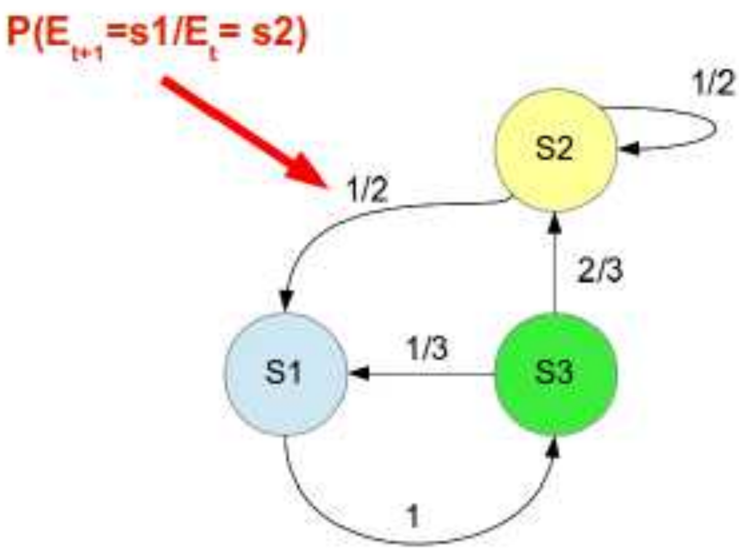


- N estados: $s_1, s_2, \ldots, s_N$
- Instantes de tiempo discretos: $t = 0, t = 1, \ldots$
- En un instante t el sistema está en un estado E_t

$$E_t \in \{s1, s2, ..., s_N\}$$

- El estado actual determina la distribución de probabilidad del estado siguiente

Modelos de Markov

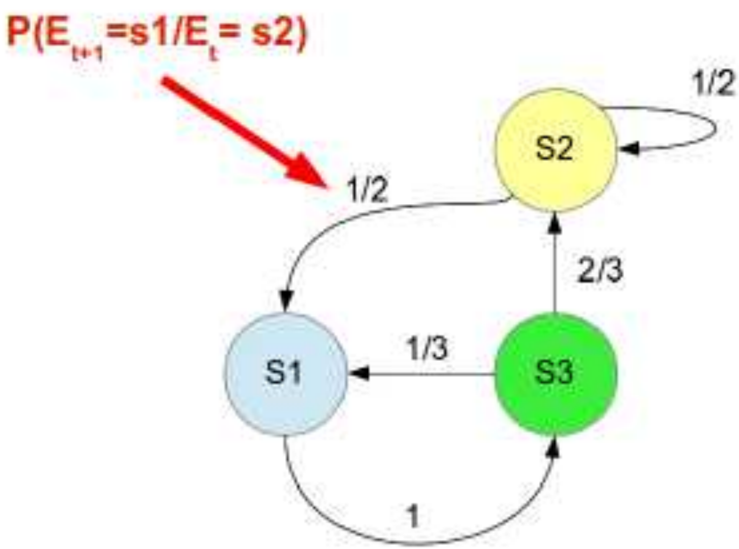


- N estados: $s_1, s_2, \ldots, s_N$
- Instantes de tiempo discretos: $t = 0, t = 1, \ldots$
- En un instante t el sistema está en un estado E_t

$$E_t \in \{s1, s2, ..., s_N\}$$

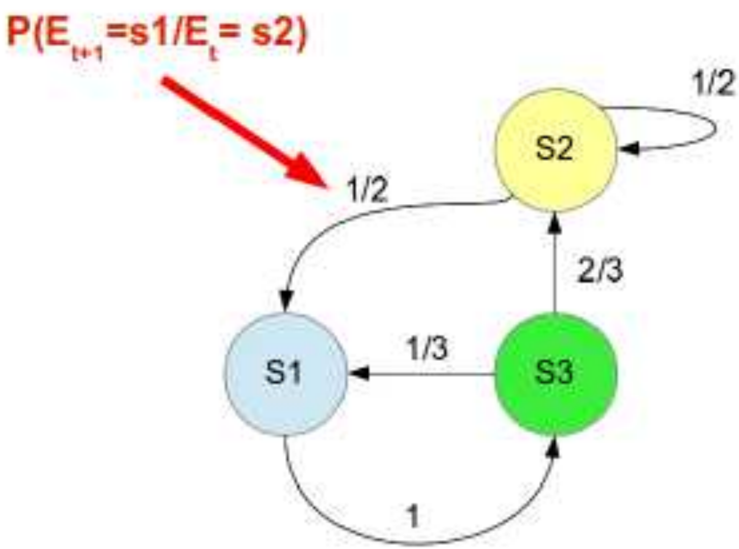
- El estado actual determina la distribución de probabilidad del estado siguiente

Modelos de Markov



- Hipótesis de Markov
  - E_{t+1} es condicionalmente independiente de $E_{t-1}, \ldots, E_0$ dado E_t
  - Es decir, el estado actual solo depende del anterior

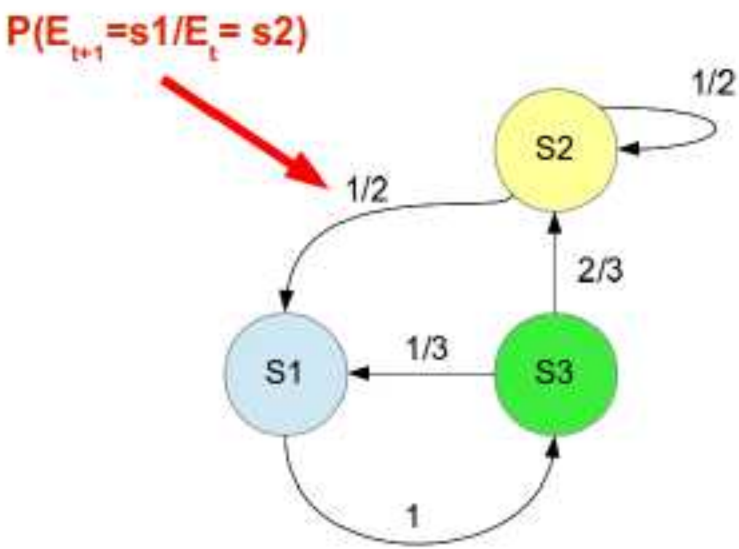
Modelos de Markov



- Hipótesis de Markov
  - E_{t+1} es condicionalmente independiente de $E_{t-1}, \ldots, E_0$ dado E_t
  - Es decir, el estado actual solo depende del anterior

$$P(E_{t+1}/E_t, E_{t-1}, \ldots) = P(E_{t+1}/E_t)$$

Modelos de Markov

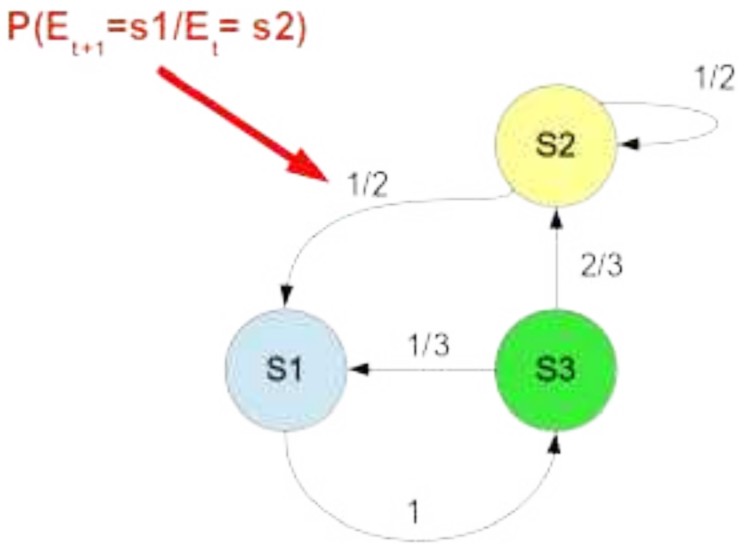


- Hipótesis de Markov
  - E_{t+1} es condicionalmente independiente de $E_{t-1}, \ldots, E_0$ dado E_t
  - Es decir, el estado actual solo depende del anterior

$$P(E_{t+1}/E_t, E_{t-1}, \ldots) = P(E_{t+1}/E_t)$$

- La distribución conjunta $P(E_0, E_1, E_2, \ldots)$ se puede representar con una red bayesiana

Modelos de Markov

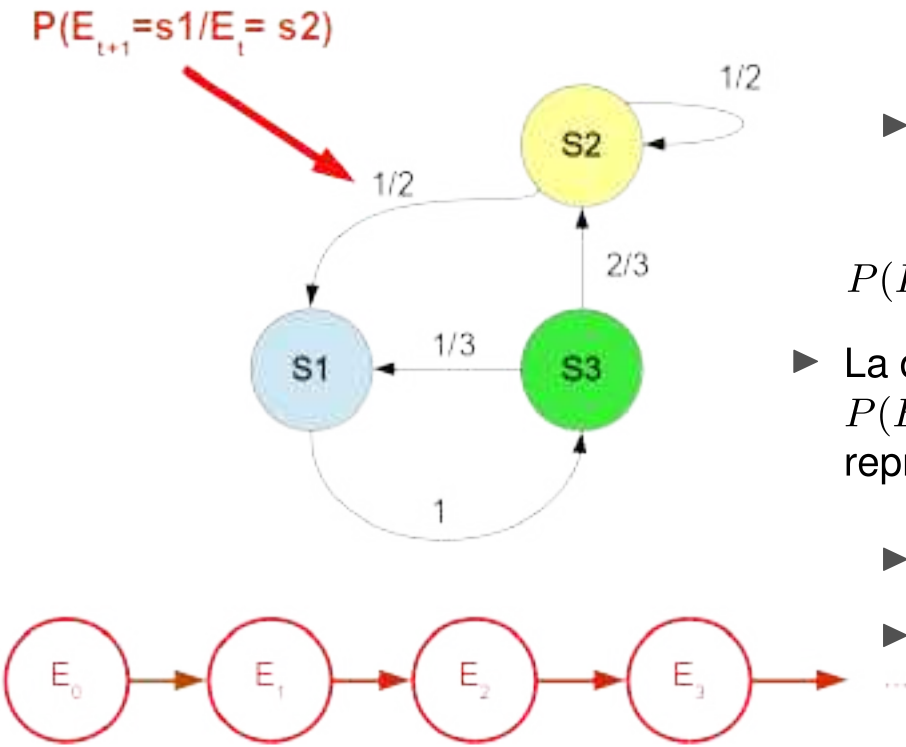


- Hipótesis de Markov
  - E_{t+1} es condicionalmente independiente de $E_{t-1}, \ldots, E_0$ dado E_t
  - Es decir, el estado actual solo depende del anterior

$$P(E_{t+1}/E_t, E_{t-1}, \ldots) = P(E_{t+1}/E_t)$$

- La distribución conjunta $P(E_0, E_1, E_2, \ldots)$ se puede representar con una red bayesiana

Modelos de Markov

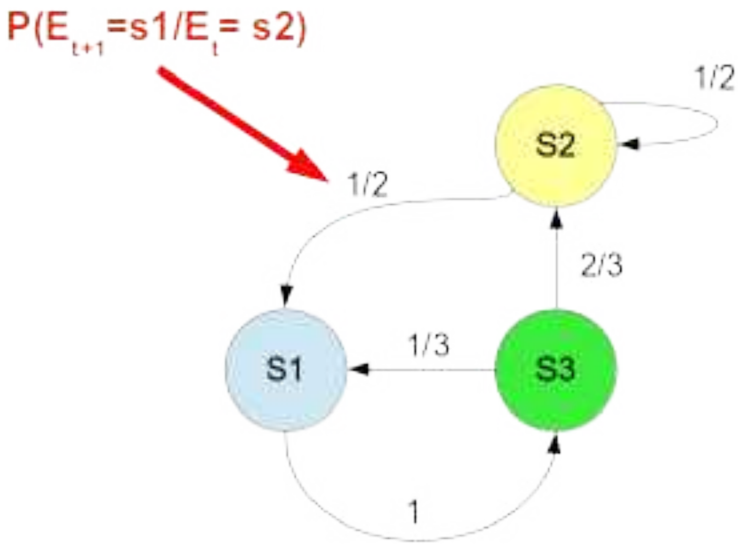


- Hipótesis de Markov
  - E_{t+1} es condicionalmente independiente de $E_{t-1}, \ldots, E_0$ dado E_t
  - Es decir, el estado actual solo depende del anterior

$$P(E_{t+1}/E_t, E_{t-1}, \ldots) = P(E_{t+1}/E_t)$$

- La distribución conjunta $P(E_0, E_1, E_2, \ldots)$ se puede representar con una red bayesiana
  - Prob a priori: $P(E_0)$
  - CPTs: $P(E_{t+1}/E_t)$

Modelos de Markov



- Hipótesis de Markov
  - E_{t+1} es condicionalmente independiente de $E_{t-1}, \ldots, E_0$ dado E_t
  - Es decir, el estado actual solo depende del anterior

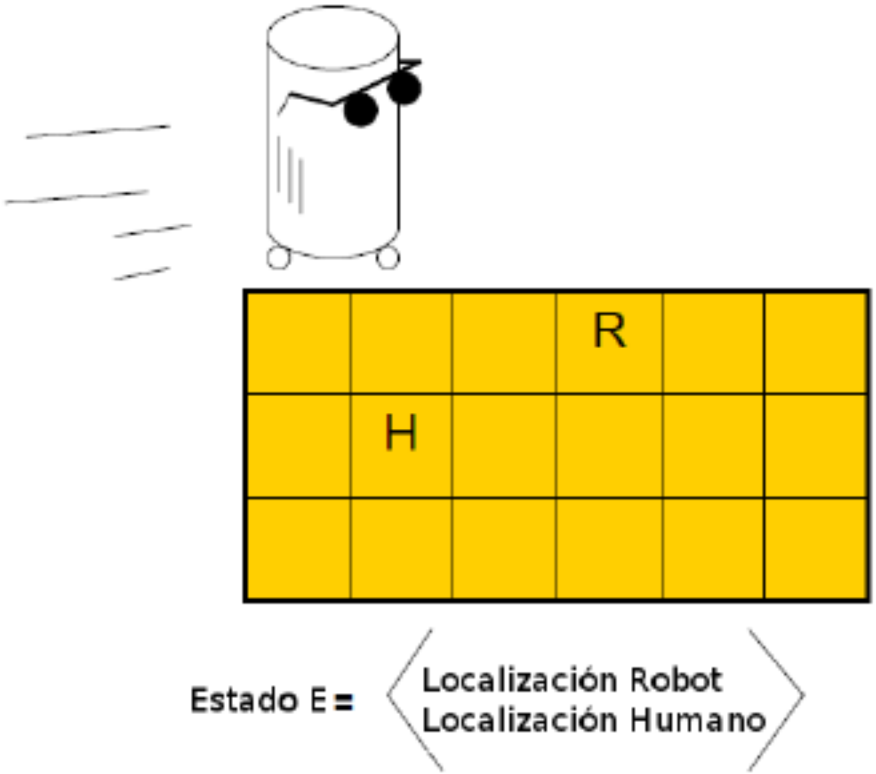
$$P(E_{t+1}/E_t, E_{t-1}, \ldots) = P(E_{t+1}/E_t)$$

- La distribución conjunta $P(E_0, E_1, E_2, \ldots)$ se puede representar con una red bayesiana
  - Prob a priori: $P(E_0)$
  - CPTs: $P(E_{t+1}/E_t)$

  ...
- Proceso estacionario: las CPTs no cambian con el tiempo $\rightarrow$ ¡basta con definir una!

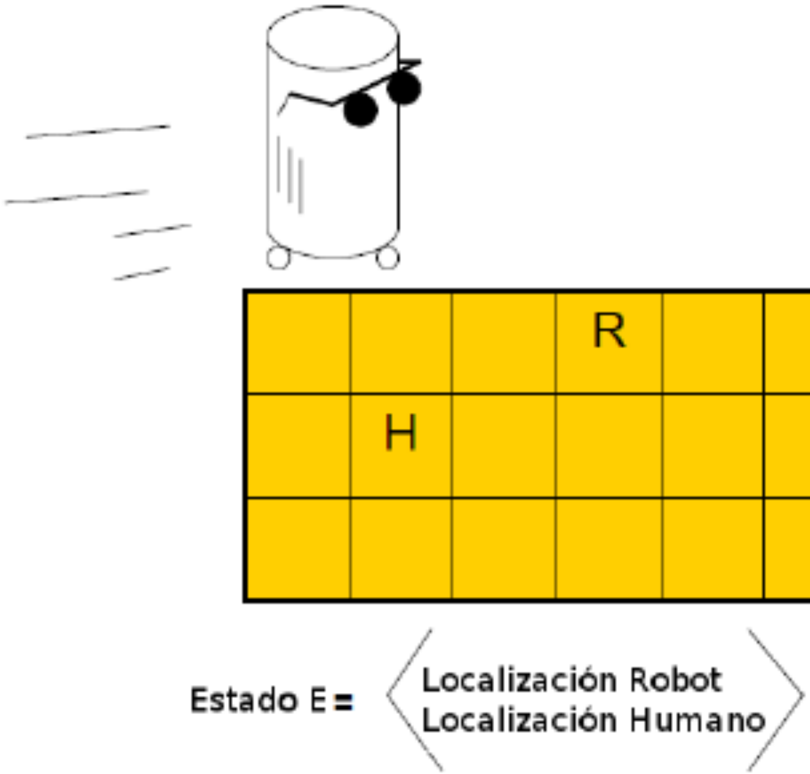
Ejemplo

Un humano y un robot se mueven por un grid ...



Ejemplo

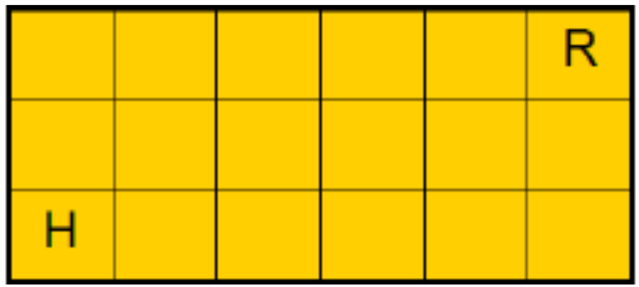
Un humano y un robot se mueven por un grid ...



Número de estados $= 18 \times 18 = 324$

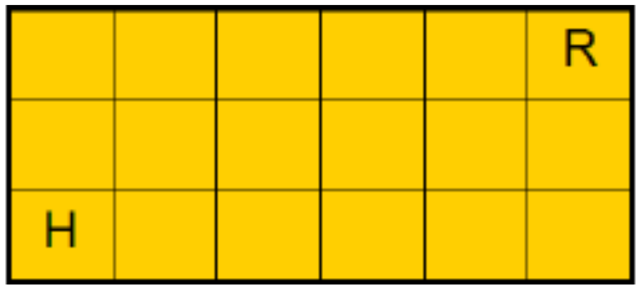
Ejemplo

- En cada instante de tiempo el robot se mueve a una celda adyacente y el humano se mueve también a una celda adyacente
- Estado inicial, E_0:



Ejemplo

- En cada instante de tiempo el robot se mueve a una celda adyacente y el humano se mueve también a una celda adyacente
- Estado inicial, E_0:



- Preguntas típicas
  - ¿Cuál es el tiempo esperado para que el robot alcance al humano?
  - ¿Cuál es la probabilidad de que el robot alcance al humano en el próximo instante?
  - ¿Cuál es la probabilidad de que el robot alcance al humano en N instantes?

Ejemplo de pregunta

Estamos en tiempo t y el humano no ha sido alcanzado. ¿Cuál es la probabilidad de que el alcance ocurra en t+1?

- Caso 1: **Si el robot es ciego** $\rightarrow$ Modelos de Markov (MM)

Ejemplo de pregunta

Estamos en tiempo t y el humano no ha sido alcanzado. ¿Cuál es la probabilidad de que el alcance ocurra en t+1?

- Caso 1: **Si el robot es ciego** $\rightarrow$ Modelos de Markov (MM)
- Caso 2: Si el robot tiene algunos sensores, pero la información sobre el estado es incompleta $\rightarrow$ Modelos de Markov Ocultos (HMM)

Ejemplo de pregunta

Estamos en tiempo t y el humano no ha sido alcanzado. ¿Cuál es la probabilidad de que el alcance ocurra en t+1?

- Caso 1: **Si el robot es ciego** $\rightarrow$ Modelos de Markov (MM) Veremos esto primero
- Caso 2: Si el robot tiene algunos sensores, pero la información sobre el estado es incompleta $\rightarrow$ Modelos de Markov Ocultos (HMM)

Modelos de Markov: ¿Cuál es la probabilidad P($E_t = s_j$)?

- Paso 1: Calcular la probabilidad de todos los posibles caminos
  - Dado que conocemos $P(E_0 = s_i) = 1$

$$P(E_0, E_1, \ldots, E_t) = P(E_1/E_0)P(E_2/E_1)\ldots P(E_t/E_{t-1})$$

Modelos de Markov: ¿Cuál es la probabilidad P($E_t = s_j$)?

- Paso 1: Calcular la probabilidad de todos los posibles caminos
  - Dado que conocemos $P(E_0 = s_i) = 1$

$$P(E_0, E_1, \ldots, E_t) = P(E_1/E_0)P(E_2/E_1)\ldots P(E_t/E_{t-1})$$

- Paso 2: marginamos $P(E_t = s_j)$

$$P(E_t = s_j) = \sum_{\mathrm{Todos\ los\ caminos\ que\ acaban\ en\ s_j}} P(E_0, E_1, \ldots, E_t = s_j)$$

Modelos de Markov: ¿Cuál es la probabilidad P($E_t = s_j$)?

- Paso 1: Calcular la probabilidad de todos los posibles caminos
  - Dado que conocemos $P(E_0 = s_i) = 1$

$$P(E_0, E_1, \ldots, E_t) = P(E_1/E_0)P(E_2/E_1)\ldots P(E_t/E_{t-1})$$

- Paso 2: marginamos $P(E_t = s_j)$

$$P(E_t = s_j) = \sum_{\text{Todos los caminos que acaban en s}_\text{j}} P(E_0, E_1, \ldots, E_t = s_j)$$

- **¡El cálculo es exponencial en t!**

¿Cuál es la probabilidad P($E_t = s_j$)? Método más inteligente: Programación dinámica

- Definir para cada estado s_i la probabilidad de estar en ese estado en el instante t

$$P(E_t = s_i)$$

¿Cuál es la probabilidad P($E_t = s_j$)? Método más inteligente: Programación dinámica

- Definir para cada estado s_i la probabilidad de estar en ese estado en el instante t

$$P(E_t = s_i)$$

- Mediante una definición recursiva:

$$\forall i, P(E_0 = s_i) = \begin{cases} 1 & \text{Si } s_i \text{ es el estado inicial} \\ 0 & \text{En otro caso} \end{cases}$$

¿Cuál es la probabilidad P($E_t = s_j$)? Método más inteligente: Programación dinámica

- Definir para cada estado s_i la probabilidad de estar en ese estado en el instante t

$$P(E_t = s_i)$$

- Mediante una definición recursiva:

$$\forall i, P(E_0 = s_i) = \begin{cases} 1 & \text{Si } s_i \text{ es el estado inicial} \\ 0 & \text{En otro caso} \end{cases}$$

$$\forall j, P(E_{t+1} = s_j) = \sum_{i=1}^{N} P(E_{t+1} = s_j, E_t = s_i) = \sum_{i=1}^{N} P(E_{t+1} = s_j / E_t = s_i) P(E_t = s_i)$$

¿Cuál es la probabilidad P($E_t = s_j$)? Método más inteligente: Programación dinámica

- Definir para cada estado s_i la probabilidad de estar en ese estado en el instante t

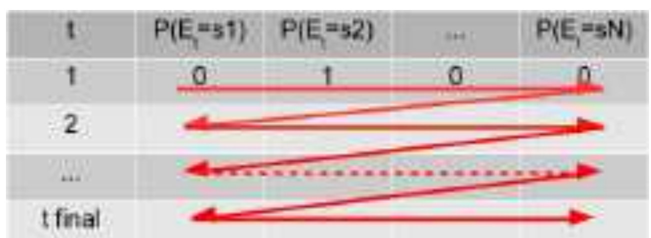
$$P(E_t = s_i)$$

- Mediante una definición recursiva:

$$\forall i, P(E_0 = s_i) = \begin{cases} 1 & \text{Si } s_i \text{ es el estado inicial} \\ 0 & \text{En otro caso} \end{cases}$$

$$\forall j, P(E_{t+1} = s_j) = \sum_{i=1}^{N} P(E_{t+1} = s_j, E_t = s_i) = \sum_{i=1}^{N} P(E_{t+1} = s_j / E_t = s_i) P(E_t = s_i)$$

- El cálculo equivale a rellenar la siguiente tabla como muestran las flechas



t	P(E_t=s1)	P(E_t=s2)	...	P(E_t=sN)
1	0	1	0	0
2				
...				
t final				

¿Cuál es la probabilidad P($E_t = s_j$)? Método más inteligente: Programación dinámica

- Definir para cada estado s_i la probabilidad de estar en ese estado en el instante t

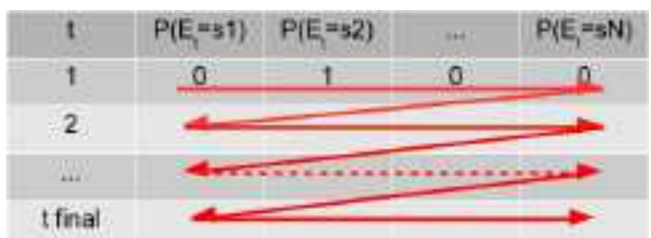
$$P(E_t = s_i)$$

- Mediante una definición recursiva:

$$\forall i, P(E_0 = s_i) = \begin{cases} 1 & \text{Si } s_i \text{ es el estado inicial} \\ 0 & \text{En otro caso} \end{cases}$$

$$\forall j, P(E_{t+1} = s_j) = \sum_{i=1}^{N} P(E_{t+1} = s_j, E_t = s_i) = \sum_{i=1}^{N} P(E_{t+1} = s_j / E_t = s_i) P(E_t = s_i)$$

- El cálculo equivale a rellenar la siguiente tabla como muestran las flechas



t	P(E$_t$=s1)	P(E$_t$=s2)	...	P(E$_t$=sN)
1	0	1	0	0
2				
...				
t final				

- Algoritmo de **Simulación hacia delante**: complejidad lineal en t

¿Cuál es la probabilidad P($E_t = s_j$)? Método más inteligente: Programación dinámica

- Definir para cada estado s_i la probabilidad de estar en ese estado en el instante t

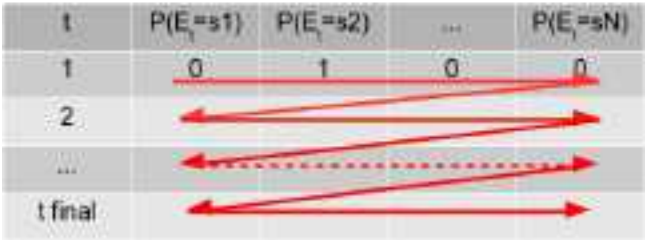
$$P(E_t = s_i)$$

- Mediante una definición recursiva:

$$\forall i, P(E_0 = s_i) = \begin{cases} 1 & \text{Si } s_i \text{ es el estado inicial} \\ 0 & \text{En otro caso} \end{cases}$$

$$\forall j, P(E_{t+1} = s_j) = \sum_{i=1}^{N} P(E_{t+1} = s_j, E_t = s_i) = \sum_{i=1}^{N} P(E_{t+1} = s_j / E_t = s_i) P(E_t = s_i)$$

- El cálculo equivale a rellenar la siguiente tabla como muestran las flechas

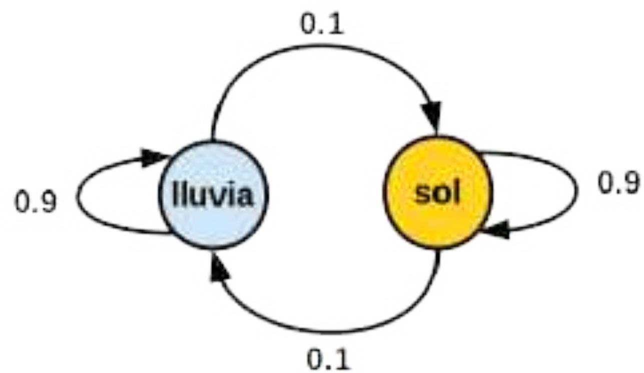


t	P(E_t=s1)	P(E_t=s2)	...	P(E_t=sN)
1	0	1	0	0
2				
...				
t final				

- Algoritmo de **Simulación hacia delante**: complejidad lineal en t

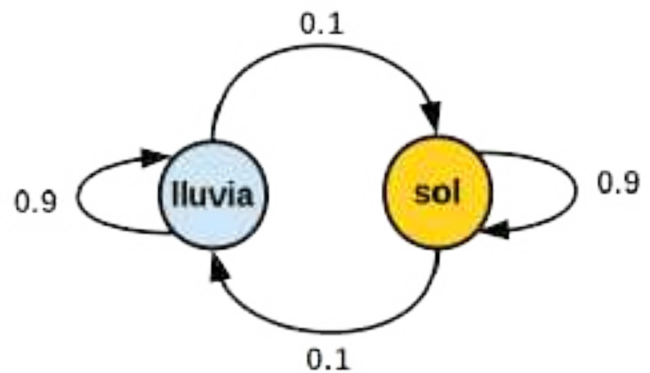
Ejemplo pequeño

- Estados: $E_t = \{Tiempo_t\}$
- Transiciones: $P(E_t/E_{t-1})$



Ejemplo pequeño

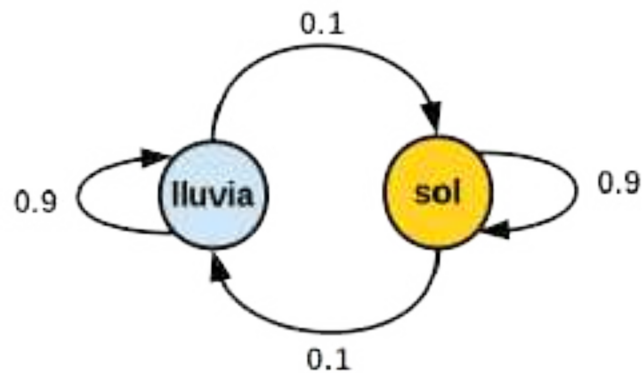
- Estados: $E_t = \{Tiempo_t\}$
- Transiciones: $P(E_t/E_{t-1})$



- Inicialmente, $P(E_0 = sol) = 1$, $P(E_0 = lluvia) = 0$

Ejemplo pequeño

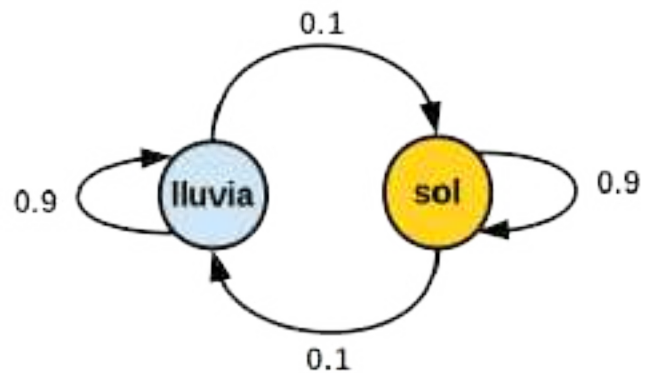
- Estados: $E_t = \{Tiempo_t\}$
- Transiciones: $P(E_t/E_{t-1})$



- Inicialmente, $P(E_0 = sol) = 1$, $P(E_0 = lluvia) = 0$
- ¿Cuál es la probabilidad de sol al día siguiente, E_1?

Ejemplo pequeño

- Estados: $E_t = \{Tiempo_t\}$
- Transiciones: $P(E_t/E_{t-1})$



- Inicialmente, $P(E_0 = sol) = 1$, $P(E_0 = lluvia) = 0$
- ¿Cuál es la probabilidad de sol al día siguiente, E_1?

$$
\begin{aligned}
P(E_1 = sol) &= P(E_1 = sol/E_0 = sol)P(E_0 = sol) + P(E_1 = sol/E_0 = lluvia)P(E_0 = lluvia) \\
&= 0.9 \times 1.0 + 0.1 \times 0.0 = 0.9
\end{aligned}
$$

Ejemplo pequeño: simulación hacia delante

- ¿Cuál es la distribución de probabilidad tras k pasos?
- Algoritmo de **simulación hacia delante**

Ejemplo pequeño: simulación hacia delante

- ¿Cuál es la distribución de probabilidad tras k pasos?
- Algoritmo de **simulación hacia delante**

$$P(E_{t+1} = s_j) = \sum_{i=1}^{N} P(E_{t+1} = s_j / E_t = s_i) P(E_t = s_i)$$

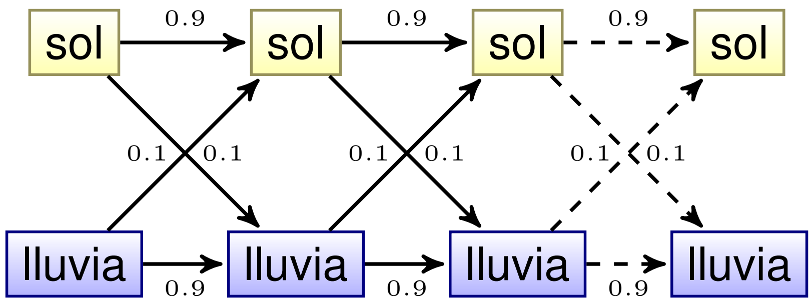
- $P(E_0)$ conocido

Ejemplo pequeño: simulación hacia delante

- ¿Cuál es la distribución de probabilidad tras k pasos?
- Algoritmo de **simulación hacia delante**

$$P(E_{t+1} = s_j) = \sum_{i=1}^{N} P(E_{t+1} = s_j / E_t = s_i) P(E_t = s_i)$$

- $P(E_0)$ conocido



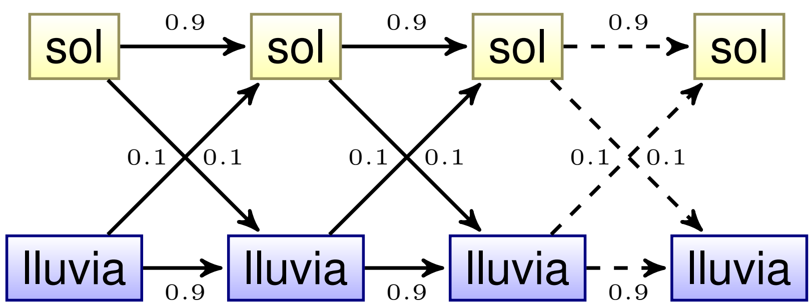
Ejemplo pequeño: simulación hacia delante

- ¿Cuál es la distribución de probabilidad tras k pasos?
- Algoritmo de **simulación hacia delante**

$$P(E_{t+1} = s_j) = \sum_{i=1}^{N} P(E_{t+1} = s_j / E_t = s_i) P(E_t = s_i)$$

- $P(E_0)$ conocido

$\boxed{P(E_t = sol) :}$



$\boxed{P(E_t = lluvia) :}$

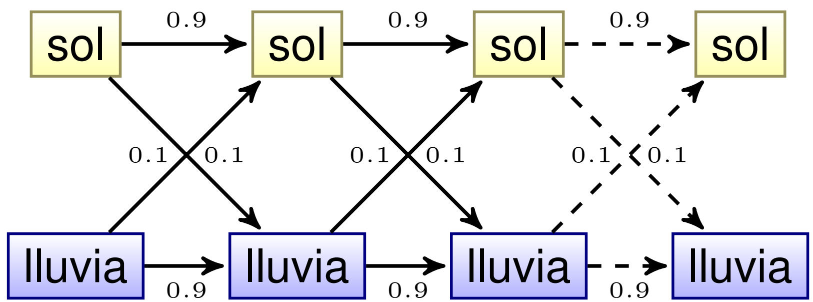
Ejemplo pequeño: simulación hacia delante

- ¿Cuál es la distribución de probabilidad tras k pasos?
- Algoritmo de **simulación hacia delante**

$$P(E_{t+1} = s_j) = \sum_{i=1}^{N} P(E_{t+1} = s_j / E_t = s_i) P(E_t = s_i)$$

- $P(E_0)$ conocido E_0

$\boxed{P(E_t = sol) :}$ $\boxed{1.0}$



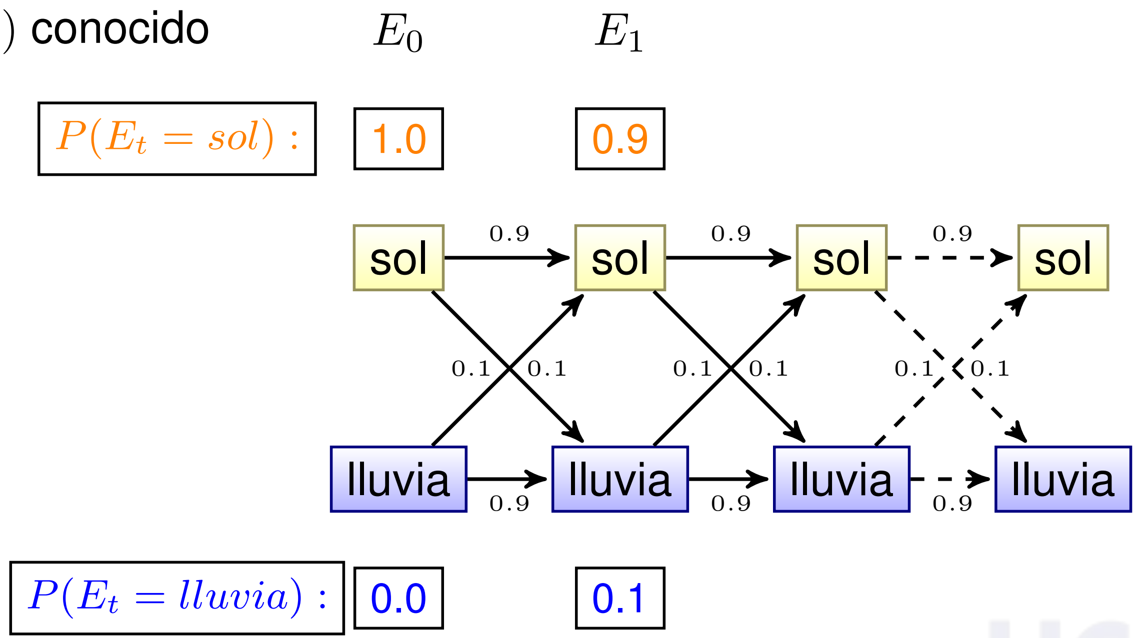
$\boxed{P(E_t = lluvia) :}$ $\boxed{0.0}$

Ejemplo pequeño: simulación hacia delante

- ¿Cuál es la distribución de probabilidad tras k pasos?
- Algoritmo de **simulación hacia delante**

$$P(E_{t+1} = s_j) = \sum_{i=1}^{N} P(E_{t+1} = s_j / E_t = s_i) P(E_t = s_i)$$

- $P(E_0)$ conocido



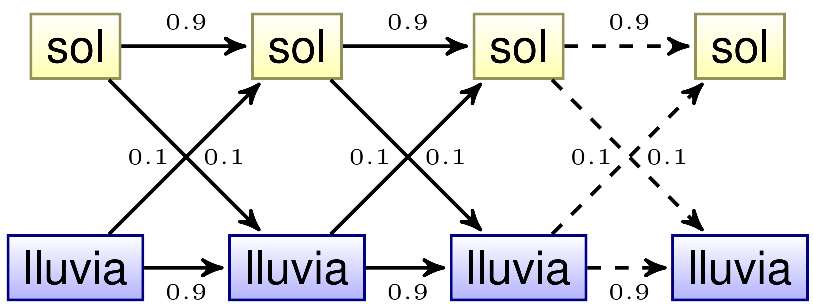
Ejemplo pequeño: simulación hacia delante

- ¿Cuál es la distribución de probabilidad tras k pasos?
- Algoritmo de **simulación hacia delante**

$$P(E_{t+1} = s_j) = \sum_{i=1}^{N} P(E_{t+1} = s_j / E_t = s_i) P(E_t = s_i)$$

- $P(E_0)$ conocido

	E_0	E_1	E_2
$P(E_t = sol)$:	1.0	0.9	0.82



	E_0	E_1	E_2
$P(E_t = lluvia)$:	0.0	0.1	0.18

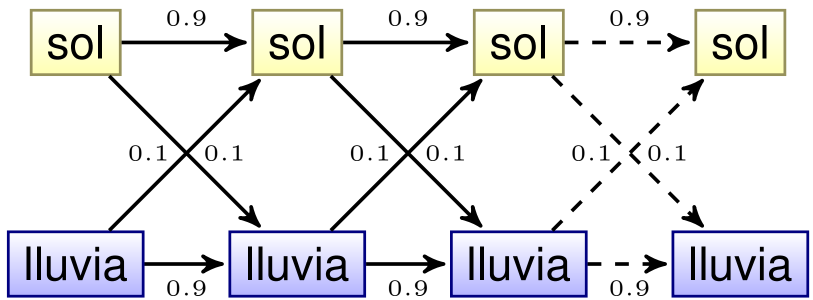
Ejemplo pequeño: simulación hacia delante

- ¿Cuál es la distribución de probabilidad tras k pasos?
- Algoritmo de **simulación hacia delante**

$$P(E_{t+1} = s_j) = \sum_{i=1}^{N} P(E_{t+1} = s_j / E_t = s_i) P(E_t = s_i)$$

- $P(E_0)$ conocido

	E_0	E_1	E_2	E_∞
$P(E_t = sol)$:	1.0	0.9	0.82	0.5
$P(E_t = lluvia)$:	0.0	0.1	0.18	0.5



Distribuciones estacionarias

- Si simulamos la cadena una cantidad suficiente de pasos
  - La incertidumbre se acumula
  - Eventualmente, ¡no sabremos cúal es el estado!
- Distribuciones estacionarias
  - Para la mayoría de las cadenas la distribución al final es independiente de la inicial
    la distribución al final es igual si empezamos con $P(X_0 = lluvia) = 1$
  - **Normalmente sólo podemos predecir a corto plazo**

En este tema

Incertidumbre

Razonamiento Probabilístico
Redes Bayesianas
Razonamiento Probabilístico en el tiempo
- Introducción
- Modelos de Markov
- Modelos de Markov Ocultos (HMMs)
- Procesos de Decisión de Markov (MDPs)
- Procesos de Decisión de Markov Parcialmente Observables (POMDPs)

Lógica Borrosa

Volviendo al ejemplo del robot

Estamos en tiempo t y el humano no ha sido alcanzado. ¿Cuál es la probabilidad de que el alcance ocurra en t+1?

- Caso 1: Si el robot es ciego $\rightarrow$ Modelos de Markov (MM) (Hemos visto ésto)

Volviendo al ejemplo del robot

Estamos en tiempo t y el humano no ha sido alcanzado. ¿Cuál es la probabilidad de que el alcance ocurra en t+1?

- Caso 1: Si el robot es ciego $\rightarrow$ Modelos de Markov (MM) (Hemos visto ésto)
- Caso 2: Si el robot tiene algunos sensores, pero la información sobre el estado es incompleta $\rightarrow$ Modelos de Markov Ocultos (HMM)

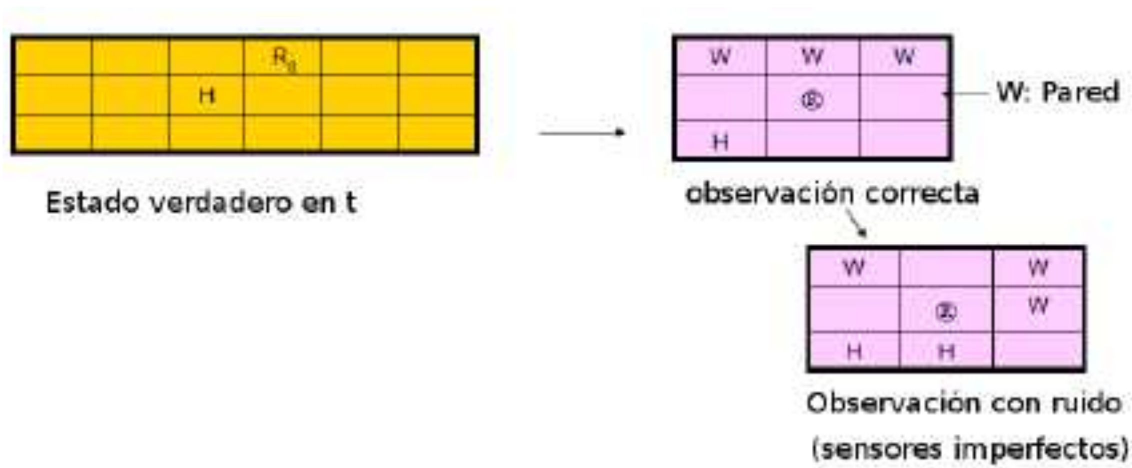
Volviendo al ejemplo del robot

Estamos en tiempo t y el humano no ha sido alcanzado. ¿Cuál es la probabilidad de que el alcance ocurra en t+1?

- Caso 1: Si el robot es ciego $\rightarrow$ Modelos de Markov (MM) (Hemos visto ésto)
- Caso 2: Si el robot tiene algunos sensores, pero la información sobre el estado es incompleta $\rightarrow$ Modelos de Markov Ocultos (HMM) Ahora veremos ésto

Modelos de Markov Ocultos (HMM)

- El estado no es observable (está oculto), pero...
- Hay variables observables (evidencia) que determinan el estado
- **Ejemplo robot: sensores de proximidad. El robot ve las 8 celdas adyacentes**
- **Los sensores son imperfectos**



Modelos de Markov Ocultos (HMMs)

- Se asume
  - Que es un proceso estacionario
  - Que se cumple la hipótesis de Markov
  - Que las observaciones en t (O_t) sólo dependen del estado en t (E_t)

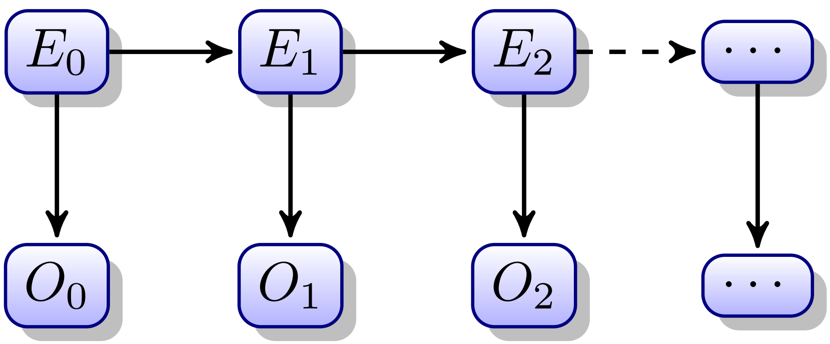
$$P(O_t/E_t, \text{historia anterior}) = P(O_t/E_t)$$

Modelos de Markov Ocultos (HMMs)

- Se asume
  - Que es un proceso estacionario
  - Que se cumple la hipótesis de Markov
  - Que las observaciones en t (O_t) sólo dependen del estado en t (E_t)

$$P(O_t/E_t, \text{historia anterior}) = P(O_t/E_t)$$

  - La red bayesiana correspondiente es:



HMMs

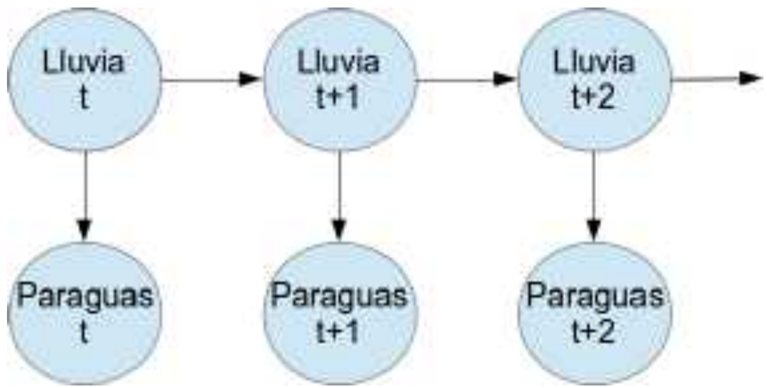
- Un HMM se define por
  - Un conjunto de estados
  - Un conjunto de observaciones
  - La probabilidad a priori $P(E_0)$
  - El modelo de transiciones $P(E_{t+1}/E_t)$
  - El modelo de observaciones $P(O_t/E_t)$

Exemplo pequeño

- *Un estudiante estudia día tras día en su estudio sin ventanas, pero ve a una persona entrar de la calle cada día y puede ver si lleva o no paraguas*
- Estados: Lluvia (si, no)
- Observaciones: Paraguas (si, no)

Exemplo pequeño

- *Un estudiante estudia día tras día en su estudio sin ventanas, pero ve a una persona entrar de la calle cada día y puede ver si lleva o no paraguas*
- Estados: Lluvia (si, no)
- Observaciones: Paraguas (si, no)



Tareas Típicas de Inferencia

- Problema de Evaluación : calcular la probabilidad de una secuencia de observaciones

$$P(O_o, O_1, \ldots, O_t)$$

¿Cuál es la probabilidad de llevar paraguas tres días seguidos?

Tareas Típicas de Inferencia

- Problema de Decodificación: dada una secuencia de observaciones $O_o, O_1, \ldots, O_t$, determinar cuál es la secuencia de estados correspondiente que *explica mejor* esas observaciones.

$$\underset{E_0,\ldots,E_{n+1}}{\arg\max}\ P(O_o, \ldots, O_n, E_1, \ldots, E_{n+1})$$

Dado que llevó paraguas tres días seguidos, ¿cuál es la probabilidad de que esos tres días haya llovido?

Tareas Típicas de Inferencia

- Problema de Filtrado. Distribución de probabilidad del estado actual dada cierta evidencia histórica

$$P(E_t/O_0, \ldots, O_t)$$

Dado que la persona llevó paraguas tres días ¿cuál es la probabilidad de que esté lloviendo el día tres?

Tareas Típicas de Inferencia

- Problema de Predicción. Distribución de probabilidad de estados futuros dada evidencia

$$P(E_{t+k}/O_0, \ldots, O_t)$$

Dado que la persona llevó paraguas tres días ¿cuál es la probabilidad de que llueva pasado mañana?

Inferencia en HMMs

- En los HMM la distribución conjunta factoriza como:

$$P(E_0, \ldots, E_t, O_0, \ldots, O_t) = P(O_0/E_0)P(E_0)\prod_{t=1}^{T} P(E_t/E_{t-1})P(O_t/E_t)$$

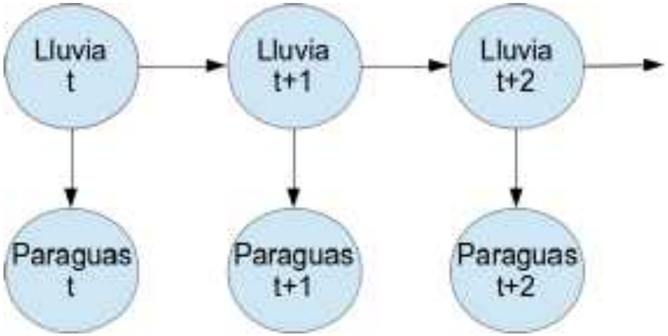
Inferencia en HMMs

- En los HMM la distribución conjunta factoriza como:

$$P(E_0, \ldots, E_t, O_0, \ldots, O_t) = P(O_0/E_0)P(E_0)\prod_{t=1}^{T} P(E_t/E_{t-1})P(O_t/E_t)$$

- Al igual que en MM es posible una definición recursiva
- Muchas de las tareas de inferencia en HMM se resuelven de forma eficiente con programación dinámica (aunque no veremos ésto en este curso)
- Resolveremos HMM mediante inferencia exacta en redes bayesianas.

Ejemplo pequeño de filtrado



$P(LLuvia_0)$: $P(Lluvia_0 = si) = 0.5$

$P(Lluvia_{t+1}/Lluvia_t)$:

L_t	$P(L_{t+1} = si/L_t)$
si	0.7
no	0.3

$P(Paraguas_t/Lluvia_t)$:

L_t	$P(P_t = si/L_t)$
si	0.9
no	0.2

Ejemplo pequeño de filtrado

- Probabilidad de que hoy esté lloviendo si ha traído paraguas dos días seguidos

$$P(L_1 = si/P_0 = si, P_1 = si)$$

Ejemplo pequeño de filtrado

- Probabilidad de que hoy esté lloviendo si ha traído paraguas dos días seguidos

$$P(L_1 = si/P_0 = si, P_1 = si)$$

$$\begin{aligned}
P(L_1 = si/P_0 = si, P_1 = si) = & \\
&= \alpha P(L_1 = si, P_0 = si, P_1 = si) \\
&= \alpha \sum_{l0} P(L_0 = l_0, L_1 = si, P_0 = si, P_1 = si) \\
&= \alpha \sum_{l0} P(L_1 = si/L_0 = l_0) P(L_0 = l_0) P(P_0 = si/L_0 = l_0) P(P_1 = si/L_1 = si) \\
&= \alpha P(P_1 = si/L_1 = si) \sum_{l0} P(L_1 = si/L_0 = l_0) P(L_0 = l_0) P(P_0 = si/L_0 = l_0) \\
&= \alpha P(P_1 = si/L_1 = si)[P(L_1 = si/L_0 = si) P(L_0 = si) P(P_0 = si/L_0 = si) \\
&\quad + P(L_1 = si/L_0 = no) P(L_0 = no) P(P_0 = si/L_0 = no)] \\
&= \alpha \times 0.9 \times [0.7 \times 0.5 \times 0.9 + 0.3 \times 0.5 \times 0.2] \\
&= \alpha \times 0.3105
\end{aligned}$$

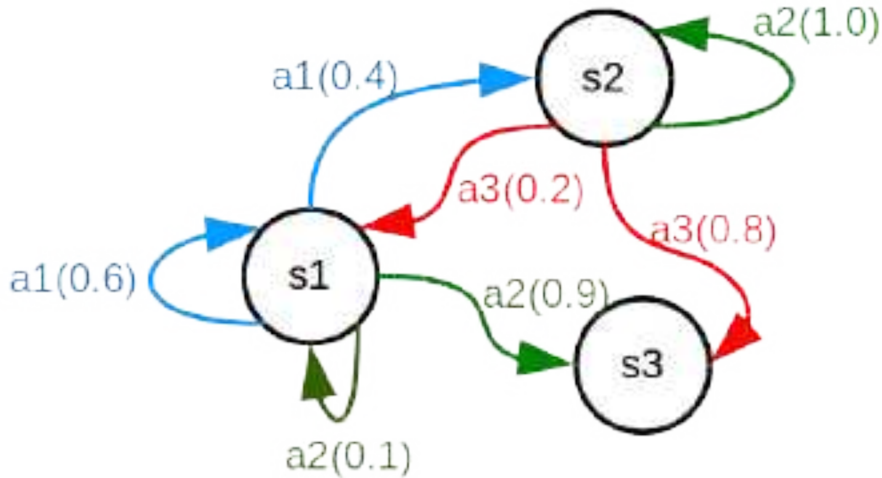
En este tema

Incertidumbre

- Razonamiento Probabilístico
- Redes Bayesianas
- Razonamiento Probabilístico en el tiempo
  - Introducción
  - Modelos de Markov
  - Modelos de Markov Ocultos (HMMs)
  - Procesos de Decisión de Markov (MDPs)
  - Procesos de Decisión de Markov Parcialmente Observables (POMDPs)
- Lógica Borrosa

Proceso de decisión de Markov (MDP)

- Proceso de Markov en el que se toma una decisión (acción) en cada instante
- Caso general: acciones no deterministas



Proceso de decisión de Markov (MDP)

- Proceso de Markov en el que se toma una decisión (acción) en cada instante

Proceso de decisión de Markov (MDP)

- Proceso de Markov en el que se toma una decisión (acción) en cada instante
- Las probabilidades de transición dependen de las acciones

$$P(S_{t+1} = s' | S_t = s, A_t = a) = P(s'/s, a)$$

Proceso de decisión de Markov (MDP)

- Proceso de Markov en el que se toma una decisión (acción) en cada instante
- Las probabilidades de transición dependen de las acciones

$$P(S_{t+1} = s' | S_t = s, A_t = a) = P(s'/s, a)$$

- Hay un refuerzo o un coste por cada par estado-acción

$$R(S_t = s, A_t = a) = R(s, a)$$

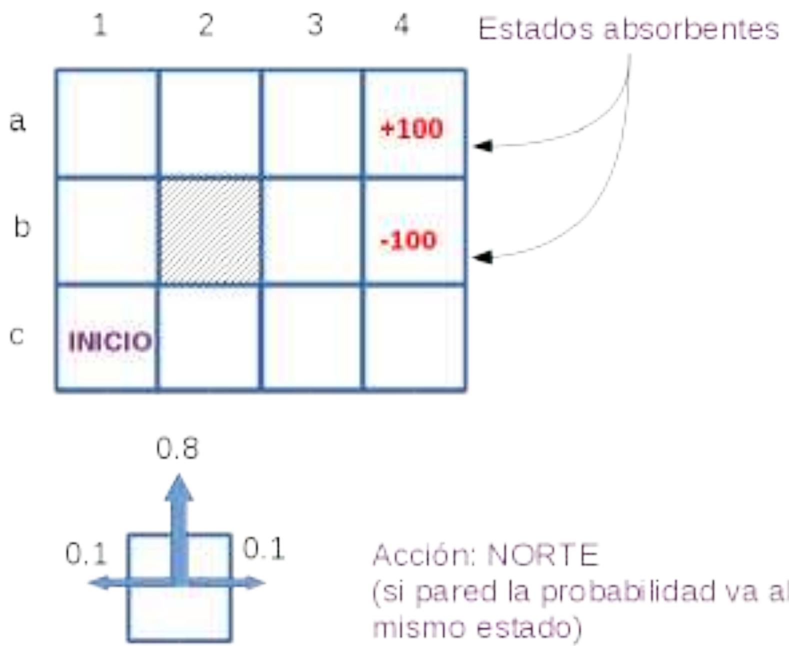
- Algunas veces se utiliza $R(s)$ en lugar de $R(s, a)$

Proceso de decisión de Markov (MDP)

- Se definen como una tupla: $< S, A, P, R >$
  - S: estados (incluyendo estado inicial y final)
  - A: acciones
  - P: tabla de transiciones $P(s'|s, a)$
  - R: refuerzo o coste $R(s, a)$
- Objetivo: determinar qué acción ejecutar en cada estado para maximizar el refuerzo o minimizar el coste a largo plazo

Ejemplo

- Acciones: norte, sur, este, oeste
- Incertidumbre sobre el resultado de la ejecución de acciones



Ejemplo: Tupla MDP: $< S, A, P, R >$

- S: posición del robot
  - Estado inicial: $X_0 = c1$
  - Estado final: $X_n = a4$ ó $b4$

Ejemplo: Tupla MDP: $< S, A, P, R >$

- S: posición del robot
  - Estado inicial: $X_0 = c1$
  - Estado final: $X_n = a4$ ó $b4$
- A: norte, sur, este, oeste

Ejemplo: Tupla MDP: $< S, A, P, R >$

- S: posición del robot
  - Estado inicial: $X_0 = c1$
  - Estado final: $X_n = a4$ ó $b4$
- A: norte, sur, este, oeste
- P: Función de transición
  - $P(X_t = b1/X_{t-1} = c1, \text{norte}) = 0.8$
  - $P(X_t = c1/X_{t-1} = c1, \text{norte}) = 0.1$
  - $P(X_t = c2/X_{t-1} = c1, \text{norte}) = 0.1$
  - ...

Ejemplo: Tupla MDP: $< S, A, P, R >$

- S: posición del robot
  - Estado inicial: $X_0 = c1$
  - Estado final: $X_n = a4$ ó $b4$
- A: norte, sur, este, oeste
- P: Función de transición
  - $P(X_t = b1/X_{t-1} = c1, \text{norte}) = 0.8$
  - $P(X_t = c1/X_{t-1} = c1, \text{norte}) = 0.1$
  - $P(X_t = c2/X_{t-1} = c1, \text{norte}) = 0.1$
  - ...
- R: refuerzo
  - $R(X = a4) = 100$, $R(X = b4) = -100$
  - Para cualquier otra casilla x: $R(X = x) = 0$

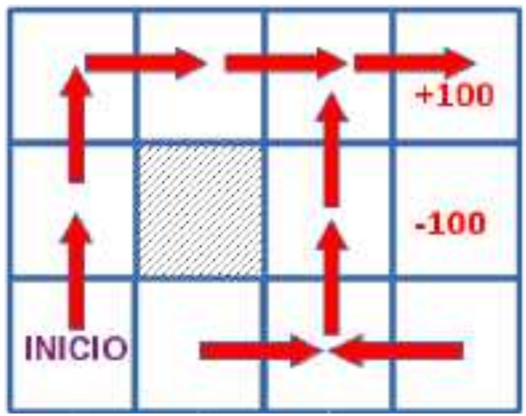
Política

- Una Política es un mapeo completo de estados a acciones
- Una política es parecido a un plan, pero no es exactamente un plan
  - Se genera hacia delante en el tiempo, como un plan
- Pero, no es una secuencia de acciones
  - Con una política, si hay fallos en la ejecución el agente puede seguir ejecutando
  - Prescribe una acción para cada estado
- Maximiza el refuerzo esperado (o minimiza el coste esperado) en lugar de alcanzar un estado meta
- Para cada MDP existe una política óptima
  - Para cada estado posible, no hay mejor opción que seguir la política

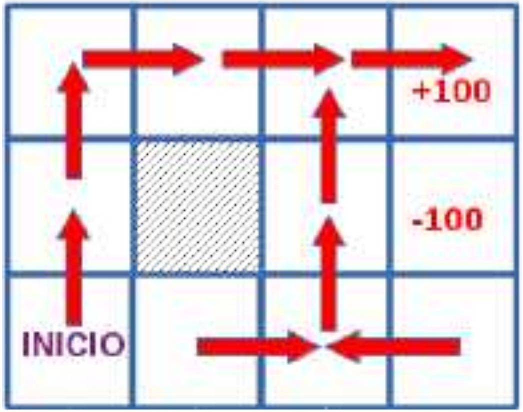
Camino (plan) vs. política (acciones estocásticas)

- Acción estocástica: consigue el efecto deseado con probabilidad p
- Ejemplo: La acción *norte* mueve hacia el norte con probabilidad 0.8, hacia el oeste con 0.1 y hacia el este con 0.1
- No hay "camino óptimo"
- La política determina qué hacer independientemente del efecto de cualquier acción en cualquier instante de tiempo

Ejemplo: robot – política

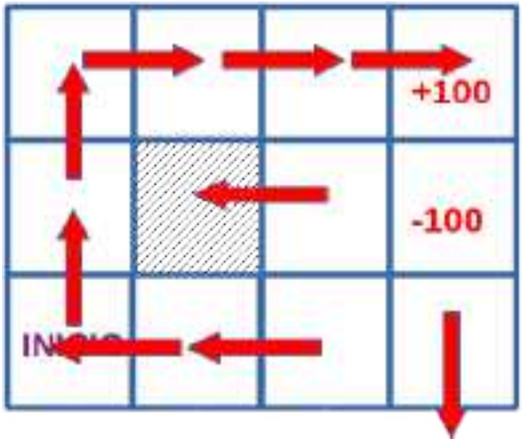


Ejemplo: robot – política

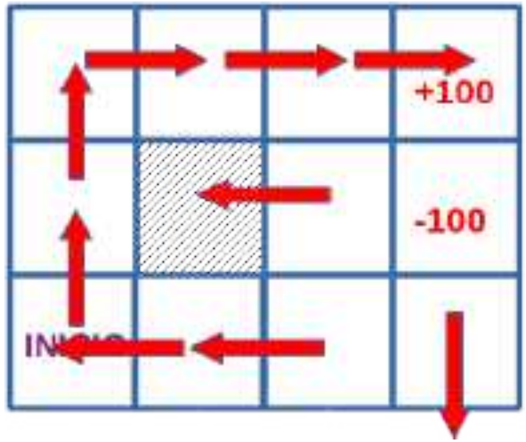


¿Cuál es la política óptima?

Ejemplo: robot – política óptima



Ejemplo: robot – política óptima



- Ésta es la política óptima porque no hay coste asociado a los demás estados
- Lo único importante es llegar al estado con refuerzo 100 evitando al máximo no caer en el estado con refuerzo -100
- No se da importancia a realizar un recorrido más largo

Ejemplo: robot – con coste (o refuerzo negativo) asociado a los demás estados

	1	2	3	4
a	-3	-3	-3	+100
b	-3		-3	-100
c	-3 INICIO	-3	-3	-3

MDP: propiedades

- Podemos razonar sobre
  - maximizar el refuerzo esperado:

$$\max E\left[\sum_{t=0}^{\infty} \gamma^t R(s_t)\right]$$

$\gamma \in [0, 1]$ es el factor de descuento

  - minimizar el coste esperado:

$$\min E\left[\sum_{t=0}^{H} C(s_t)\right]$$

Para el caso de minimizar el coste

- Política óptima π^*: mínimo coste esperado

Para el caso de minimizar el coste

- Política óptima π^*: mínimo coste esperado
  - **Acciones deterministas**: si la acción a lleva al estado s', el coste es:

$$c(a) + \text{costeDesde}(s')$$

Para el caso de minimizar el coste

- Política óptima π^*: mínimo coste esperado
  - **Acciones deterministas**: si la acción a lleva al estado s', el coste es:

$$c(a) + \mathrm{costeDesde}(s')$$

  - **Acciones no deterministas**: si la acción a tiene efectos probabílisticos, el coste esperado es:

$$c(a) + \sum_{s'} P(s'|s,a) \times \mathrm{costeDesde(s')}$$

Para el caso de minimizar el coste

- Política óptima π^*: mínimo coste esperado
  - **Acciones deterministas**: si la acción a lleva al estado s', el coste es:

$$c(a) + \mathrm{costeDesde}(s')$$

  - **Acciones no deterministas**: si la acción a tiene efectos probabílisticos, el coste esperado es:

$$c(a) + \sum_{s'} P(s'|s,a) \times \mathrm{costeDesde}(\mathrm{s}')$$

¿Cómo definimos una política óptima?

Función de valor: coste esperado de un estado

V(s**)**: coste esperado de alcanzar la meta desde s

Función de valor: coste esperado de un estado

V(s): coste esperado de alcanzar la meta desde s

- Búsqueda:
  - $V(s)$: coste del camino óptimo desde s
  - Si conocemos $V(s)$, podemos utilizar $h(s) = V(s)$ con búsqueda en escalada
  - $V(s)$ es la heurística perfecta $h^*(s)$

Función de valor: coste esperado de un estado

V(s): coste esperado de alcanzar la meta desde s

- Búsqueda:
  - $V(s)$: coste del camino óptimo desde s
  - Si conocemos $V(s)$, podemos utilizar $h(s) = V(s)$ con búsqueda en escalada
  - $V(s)$ es la heurística perfecta $h^*(s)$
- MDP
  - $V(s)$: coste esperado de la estrategia óptima para alzanzar la meta desde s
  - Si conocemos $V(s)$, podemos calcular una política óptima π^*

Función de valor: coste esperado de un estado

V(s): coste esperado de alcanzar la meta desde s

- Búsqueda:
  - $V(s)$: coste del camino óptimo desde s
  - Si conocemos $V(s)$, podemos utilizar $h(s) = V(s)$ con búsqueda en escalada
  - $V(s)$ es la heurística perfecta $h^*(s)$
- MDP
  - $V(s)$: coste esperado de la estrategia óptima para alzanzar la meta desde s
  - Si conocemos $V(s)$, podemos calcular una política óptima π^*

¿Cómo calculamos $V(S)$?

Calculando $V(S)$. Acciones deterministas

- Ecuaciones de Bellman
  - Si s es un estado meta, $V(s) = 0$
  - En otro caso (s' es el estado que se genera aplicando a en s):

$$V(s) = \min_{a \in A(s)} [c(a) + V(s')]$$

  - $A(s)$: acciones aplicables en s
- $V(s)$ es una función heurística perfecta $h^*(s)$

Ecuaciones de Bellman para MDPs (acciones no deterministas)

- Dominios estocásticos:
  - El valor esperado a partir de la acción a:

$$c(a) + \sum_{s' \in S} P_a(s'|s)V(s')$$

Ecuaciones de Bellman para MDPs (acciones no deterministas)

- Dominios estocásticos:
  - El valor esperado a partir de la acción a:

$$c(a) + \sum_{s' \in S} P_a(s'|s)V(s')$$

- Entonces,
  - Si s es un estado meta, $V(s) = 0$
  - En otro caso

$$V(s) = \min_{a \in A(s)} [c(a) + \sum_{s' \in S} P_a(s'|s)V(s')]$$

Ecuaciones de Bellman para MDPs (acciones no deterministas)

- Dominios estocásticos:
  - El valor esperado a partir de la acción a:

$$c(a) + \sum_{s' \in S} P_a(s'|s)V(s')$$

- Entonces,
  - Si s es un estado meta, $V(s) = 0$
  - En otro caso

$$V(s) = \min_{a \in A(s)} [c(a) + \sum_{s' \in S} P_a(s'|s)V(s')]$$

- Política óptima

$$\pi^*(s) = \arg\min_{a} [c(a) + \sum_{s' \in S} P_a(s'|s)V^*(s')]$$

Resolviendo las ecuaciones de Bellman

Algoritmo Iteración de Valor (Value Iteration)

$V(s_g) = 0$ para todos los estados meta s_g y todas las iteraciones
Inicializar $V_0(s) = 0$ para el resto de estados /* iteración i=0 */
while not(Fin) **do**
    for each estado $s \in S$
        /* iteración i+1 */
        $V_{i+1}(s) := \min_{a \in A(s)}[c(a) + \sum_{s' \in S} P_a(s'|s)V_i(s')]$
        Política $\pi_{i+1} = \underset{a \in A(s)}{\arg} \min[c(a) + \sum_{s' \in S} P_a(s'|s)V_i(s')]$
return π_{last}

Cuando el algoritmo se estabiliza, $\pi^* = \pi_{last}$, es la política óptima

Terminación de Value Iteration

- *Value Iteration* termina cuando alcanza un *punto fijo*, es decir, cuando los valores no cambian en dos iteraciones sucesivas
- En la práctica termina cuando $\max_s \mid V_i(s) - V_{i+1}(s) \mid \leq \epsilon$

Para el caso de refuerzo

- Se introduce un factor de descuento $\gamma \in [0, 1]$

$$\max E \left[\sum_{t=0}^{\infty} \gamma^t R(s_t) \right]$$

- El algoritmo Iteración de Valor es similar, con los siguientes cambios:
  - $V(s_g) = R(s_g)$ para todos los estados meta s_g y todas las iteraciones
  - Maximizar: $V(s) := R(s) + \max_{a \in A(s)} [\gamma \sum_{s' \in S} P_a(s'|s) V(s')]$
  - Política óptima

$$\pi^*(s) = \arg\max_a [\gamma \sum_{s' \in S} P_a(s'|s) V^*(s')]$$

Ejemplo *value iteration*: acciones deterministas

Para este ejemplo y los siguiente utlizaremos $\gamma = 1$

	1	2	3	4
a	-3	-3	-3	+100
b	-3		-3	-100
c	-3 INICIO	-3	-3	-3

$$V_{i+1}(s) := R(s) + \max_{a \in A(s)} [\gamma \sum_{s' \in S} P(s'|s,a) V_i(s')]$$

La imagen de abajo a la izquierda muestra $V_0(s)$ para todos los estados (inicialización)

	1	2	3	4
a	0	0	0	+100
b	0		0	-100
c	0	0	0	0

Ejemplo *value iteration*: acciones deterministas

Para este ejemplo y los siguiente utlizaremos $\gamma = 1$

	1	2	3	4
a	-3	-3	-3	+100
b	-3		-3	-100
c	-3 INICIO	-3	-3	-3

$$V_{i+1}(s) := R(s) + \max_{a \in A(s)} [\gamma \sum_{s' \in S} P(s'|s,a) V_i(s')]$$

La imagen de abajo a la izquierda muestra $V_0(s)$ para todos los estados (inicialización)

	1	2	3	4
a	0	0	0	+100
b	0		0	-100
c	0	0	0	0

Iteración i=1 (calcular $V_1(s)$ para todos los estados)

$V_1(a3) = -3 + \max\{0, 0, 100, 0\} = 97$

acciones en el máximo: NORTE, SUR, ESTE, OESTE
acción argmax: ESTE

Ejemplo *value iteration*: acciones deterministas

	1	2	3	4
a	-3	-3	-3	+100
b	-3		-3	-100
c	-3 INICIO	-3	-3	-3

$$V_{i+1}(s) := R(s) + \max_{a \in A(s)} [\gamma \sum_{s' \in S} P(s'|s,a) V_i(s')]$$

La imagen de abajo a la izquierda muestra $V_1(s)$ para todos los estados (iteración 1)

	1	2	3	4
a	-3	-3	97	+100
b	-3		-3	-100
c	-3	-3	-3	-3

Iteración i=2 (calcular $V_2(s)$ para todos los estados)

$V_2(b3) = -3 + \max\{97, -3, -100\} = 94$

acciones en el máximo: NORTE, SUR, ESTE
OESTE no aplicable

acción argmax: NORTE

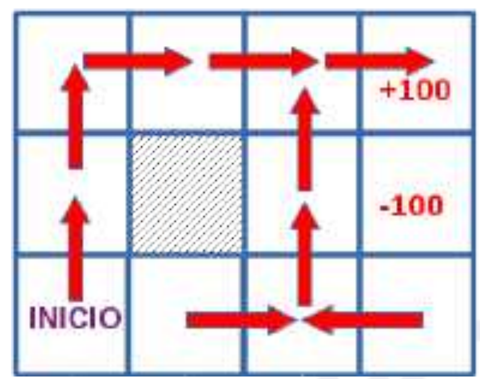
Ejemplo *value iteration*: acciones deterministas

	1	2	3	4
a	-3	-3	-3	+100
b	-3		-3	-100
c	-3 INICIO	-3	-3	-3

$$V_{i+1}(s) := R(s) + \max_{a \in A(s)} [\gamma \sum_{s' \in S} P(s'|s,a) V_i(s')]$$

La imagen de abajo a la izquierda muestra $V_{last}(s)$ para todos los estados (iteración final) y la política óptima (derecha)

	1	2	3	4
a	91	94	97	+100
b	88		94	-100
c	85	88	91	88



Ejemplo *value iteration*: acciones estocásticas

	1	2	3	4
a	-3	-3	-3	+100
b	-3		-3	-100
c	-3 INICIO	-3	-3	-3

$$V_{i+1}(s) := R(s) + \max_{a \in A(s)} [\gamma \sum_{s' \in S} P(s'|s,a) V_i(s')]$$

La imagen de abajo a la izquierda muestra $V_0(s)$ para todos los estados (iteración 0)

	1	2	3	4
a	0	0	0	+100
b	0		0	-100
c	0	0	0	0

Ejemplo *value iteration*: acciones estocásticas

	1	2	3	4
a	-3	-3	-3	+100
b	-3		-3	-100
c	-3 INICIO	-3	-3	-3

$$V_{i+1}(s) := R(s) + \max_{a \in A(s)} [\gamma \sum_{s' \in S} P(s'|s,a)V_i(s')]$$

La imagen de abajo a la izquierda muestra $V_0(s)$ para todos los estados (iteración 0)

	1	2	3	4
a	0	0	0	+100
b	0		0	-100
c	0	0	0	0

$V_1(a3)$?

Ejemplo *value iteration*: acciones estocásticas

	1	2	3	4
a	-3	-3	-3	+100
b	-3		-3	-100
c	-3 INICIO	-3	-3	-3

$$V_{i+1}(s) := R(s) + \max_{a \in A(s)} [\gamma \sum_{s' \in S} P(s'|s,a) V_i(s')]$$

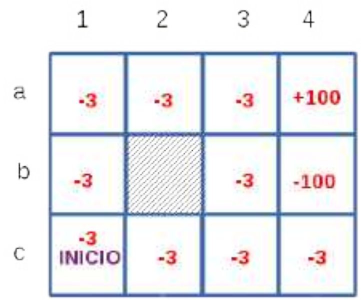
La imagen de abajo a la izquierda muestra $V_0(s)$ para todos los estados (iteración 0)

	1	2	3	4
a	0	0	0	+100
b	0		0	-100
c	0	0	0	0

$V_1(a3)$?

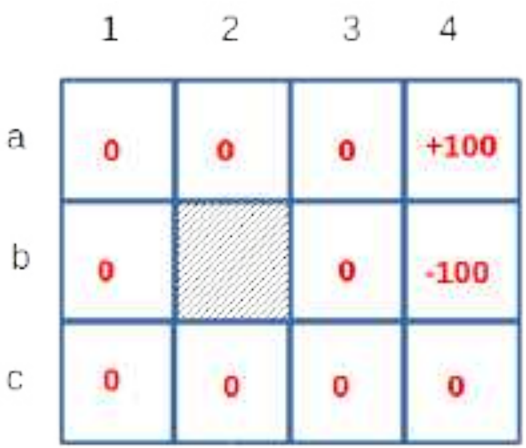
- norte $\rightarrow 0.8 \times 0 + 0.1 \times 0 + 0.1 \times 100 = 10$

Ejemplo *value iteration*: acciones estocásticas



$$V_{i+1}(s) := R(s) + \max_{a \in A(s)} [\gamma \sum_{s' \in S} P(s'|s,a) V_i(s')]$$

La imagen de abajo a la izquierda muestra $V_0(s)$ para todos los estados (iteración 0)



$V_1(a3)$?

- norte $\rightarrow 0.8 \times 0 + 0.1 \times 0 + 0.1 \times 100 = 10$
- sur $\rightarrow 0.8 \times 0 + 0.1 \times 0 + 0.1 \times 100 = 10$

Ejemplo *value iteration*: acciones estocásticas

	1	2	3	4
a	-3	-3	-3	+100
b	-3		-3	-100
c	-3 INICIO	-3	-3	-3

$$V_{i+1}(s) := R(s) + \max_{a \in A(s)} [\gamma \sum_{s' \in S} P(s'|s,a)V_i(s')]$$

La imagen de abajo a la izquierda muestra $V_0(s)$ para todos los estados (iteración 0)

	1	2	3	4
a	0	0	0	+100
b	0		0	-100
c	0	0	0	0

$V_1(a3)$?

- norte $\rightarrow 0.8 \times 0 + 0.1 \times 0 + 0.1 \times 100 = 10$
- sur $\rightarrow 0.8 \times 0 + 0.1 \times 0 + 0.1 \times 100 = 10$
- este $\rightarrow 0.8 \times 100 + 0.1 \times 0 + 0.1 \times 0 = 80$

Ejemplo *value iteration*: acciones estocásticas

	1	2	3	4
a	-3	-3	-3	+100
b	-3		-3	-100
c	-3 INICIO	-3	-3	-3

$$V_{i+1}(s) := R(s) + \max_{a \in A(s)} [\gamma \sum_{s' \in S} P(s'|s,a) V_i(s')]$$

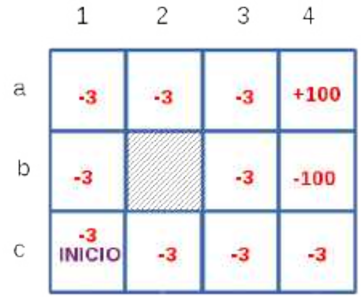
La imagen de abajo a la izquierda muestra $V_0(s)$ para todos los estados (iteración 0)

	1	2	3	4
a	0	0	0	+100
b	0		0	-100
c	0	0	0	0

$V_1(a3)$?

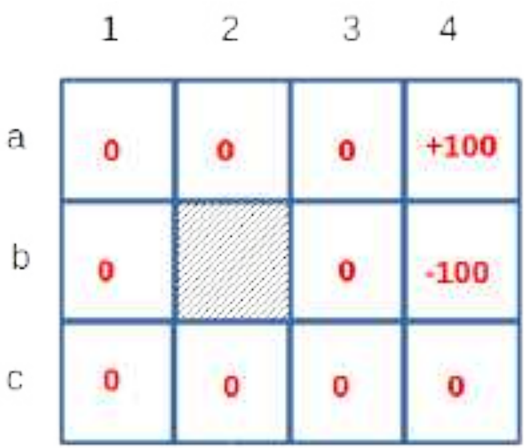
- norte $\rightarrow 0.8 \times 0 + 0.1 \times 0 + 0.1 \times 100 = 10$
- sur $\rightarrow 0.8 \times 0 + 0.1 \times 0 + 0.1 \times 100 = 10$
- este $\rightarrow 0.8 \times 100 + 0.1 \times 0 + 0.1 \times 0 = 80$
- oeste $\rightarrow 0.8 \times 0 + 0.1 \times 0 + 0.1 \times 0 = 0$

Ejemplo *value iteration*: acciones estocásticas



$$V_{i+1}(s) := R(s) + \max_{a \in A(s)} [\gamma \sum_{s' \in S} P(s'|s,a) V_i(s')]$$

La imagen de abajo a la izquierda muestra $V_0(s)$ para todos los estados (iteración 0)



$V_1(a3)$?

- norte $\rightarrow 0.8 \times 0 + 0.1 \times 0 + 0.1 \times 100 = 10$
- sur $\rightarrow 0.8 \times 0 + 0.1 \times 0 + 0.1 \times 100 = 10$
- este $\rightarrow 0.8 \times 100 + 0.1 \times 0 + 0.1 \times 0 = 80$
- oeste $\rightarrow 0.8 \times 0 + 0.1 \times 0 + 0.1 \times 0 = 0$

$V_1(a3) = -3 + \max\{10, 10, 80, 0\} = 77$
acción argmax: ESTE

Ejemplo *value iteration*: acciones estocásticas

	1	2	3	4
a	-3	-3	-3	+100
b	-3		-3	-100
c	-3 INICIO	-3	-3	-3

$$V_{i+1}(s) := R(s) + \max_{a \in A(s)} [\gamma \sum_{s' \in S} P(s'|s,a) V_i(s')]$$

La imagen de abajo a la izquierda muestra $V_1(s)$ para todos los estados (iteración 1)

	1	2	3	4
a	-3	-3	77	+100
b	-3		-3	-100
c	-3	-3	-3	-3

Ejemplo *value iteration*: acciones estocásticas

	1	2	3	4
a	-3	-3	-3	+100
b	-3		-3	-100
c	-3 INICIO	-3	-3	-3

$$V_{i+1}(s) := R(s) + \max_{a \in A(s)} [\gamma \sum_{s' \in S} P(s'|s,a) V_i(s')]$$

La imagen de abajo a la izquierda muestra $V_1(s)$ para todos los estados (iteración 1)

	1	2	3	4
a	-3	-3	77	+100
b	-3		-3	-100
c	-3	-3	-3	-3

$V_2(b3)$?

- norte $\rightarrow 0.8 \times 77 + 0.1 \times 0 + 0.1 \times -100 = 51.6$
- sur $\rightarrow 0.8 \times 0 + 0.1 \times 0 + 0.1 \times -100 = -10$
- este $\rightarrow 0.8 \times -100 + 0.1 \times 77 + 0.1 \times 0 = -72.3$
- oeste $\rightarrow 0.8 \times 0 + 0.1 \times 77 + 0.1 \times 0 = 7.7$

$V_2(b3) = -3 + max\{51.6, -10, -72.3, 7.7\} = 48.6$
acción argmax: NORTE

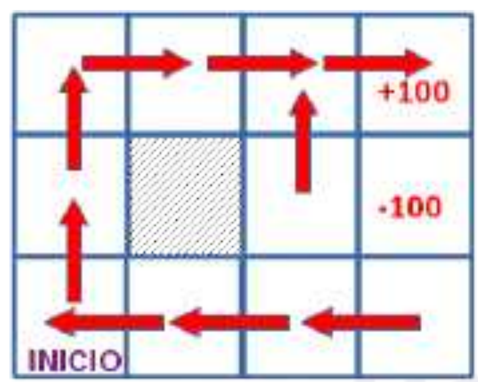
Ejemplo *value iteration*: acciones estocásticas

	1	2	3	4
a	-3	-3	-3	+100
b	-3		-3	-100
c	-3 INICIO	-3	-3	-3

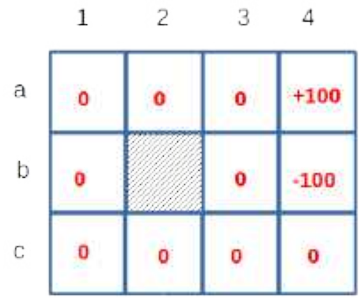
$$V_{i+1}(s) := R(s) + \max_{a \in A(s)} [\gamma \sum_{s' \in S} P(s'|s,a)V_i(s')]$$

Resultado de la última iteración y política óptima

	1	2	3	4
a	85	89	93	+100
b	81		68	-100
c	77	73	70	47



Ejemplo *value iteration*: acciones estocásticas (Refuerzos 0)



$$V_{i+1}(s) := R(s) + \max_{a \in A(s)} [\gamma \sum_{s' \in S} P(s'|s,a) V_i(s')]$$

Resultado de la última iteración y política óptima

Ejemplo *value iteration*: acciones estocásticas (Refuerzos 0)

	1	2	3	4
a	0	0	0	+100
b	0		0	-100
c	0	0	0	0

$$V_{i+1}(s) := R(s) + \max_{a \in A(s)} [\gamma \sum_{s' \in S} P(s'|s,a) V_i(s')]$$

Resultado de la última iteración y política óptima

	1	2	3	4
a	100	100	100	+100
b	100		100	-100
c	100	100	100	100

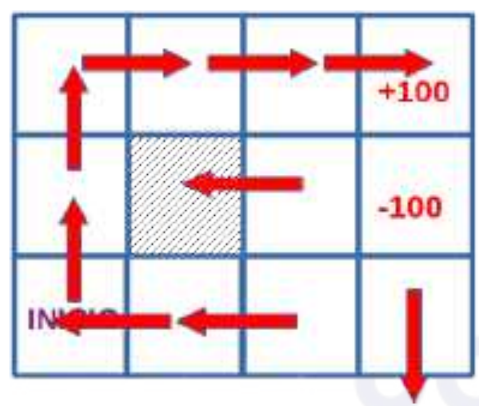
Ejemplo *value iteration*: acciones estocásticas (Refuerzos 0)

	1	2	3	4
a	0	0	0	+100
b	0		0	-100
c	0	0	0	0

$$V_{i+1}(s) := R(s) + \max_{a \in A(s)} [\gamma \sum_{s' \in S} P(s'|s,a) V_i(s')]$$

Resultado de la última iteración y política óptima

	1	2	3	4
a	100	100	100	+100
b	100		100	-100
c	100	100	100	100



Ejemplo de Value Iteration

- Estados: s_1, s_2, s_3
- Meta: s_3
- Acciones
  - Acción **Derecha**
    - s_i pasa a s_{i+1} con probabilidad 0.8
    - con probabilidad 0.2, permanece en s_i
    - coste: 1
  - Acción **Izquierda**
    - s_i pasa a s_{i-1} con probabilidad 0.8
    - con probabilidad 0.2, permanece en s_i
    - coste: 1
- Inicialmente: $V(s_1) = V(s_2) = 0$
- $V(s_3) = 0$ siempre porque es la meta

Ejemplo de Value Iteration

Iteración 1: $V(s_1) = c(\text{Derecha}) + (0.8V(s_2) + 0.2V(s_1)) = 1$
$V(s_2) = \min\{c(\text{Derecha}) + (0.8V(s_3) + 0.2V(s_2)), c(\text{Izquierda}) + (0.8V(s_1) + 0.2V(s_2)) = \min\{1, 1\} = 1$

Ejemplo de Value Iteration

Iteración 1: $V(s_1) = c(\text{Derecha}) + (0.8V(s_2) + 0.2V(s_1)) = 1$
$V(s_2) = \min\{c(\text{Derecha}) + (0.8V(s_3) + 0.2V(s_2)), c(\text{Izquierda}) + (0.8V(s_1) + 0.2V(s_2)) = \min\{1, 1\} = 1$

Iteración 2: $V(s_1) = 1 + (0.8 \times 1 + 0.2 \times 1) = 2$
$V(s_2) = \min\{1 + (0.8 \times 0 + 0.2 \times 1), 1 + (0.8 \times 1 + 0.2 \times 1)\} = \min\{1.2, 2\} = 1.2$

Ejemplo de Value Iteration

Iteración 1: $V(s_1) = c(\text{Derecha}) + (0.8V(s_2) + 0.2V(s_1)) = 1$
$V(s_2) = \min\{c(\text{Derecha}) + (0.8V(s_3) + 0.2V(s_2)), c(\text{Izquierda}) + (0.8V(s_1) + 0.2V(s_2)) = \min\{1, 1\} = 1$

Iteración 2: $V(s_1) = 1 + (0.8 \times 1 + 0.2 \times 1) = 2$
$V(s_2) = \min\{1 + (0.8 \times 0 + 0.2 \times 1), 1 + (0.8 \times 1 + 0.2 \times 1)\} = \min\{1.2, 2\} = 1.2$

Iteración 3: $V(s_1) = 1 + (0.8 \times 1.2 + 0.2 \times 2) = 2.36$
$V(s_2) = \min\{1 + (0.8 \times 0 + 0.2 \times 1.2), 1 + (0.8 \times 2 + 0.2 \times 1.2)\} = \min\{1.24, 2.84\}$

Ejemplo de Value Iteration

Iteración 1: $V(s_1) = c(\text{Derecha}) + (0.8V(s_2) + 0.2V(s_1)) = 1$
$V(s_2) = \min\{c(\text{Derecha}) + (0.8V(s_3) + 0.2V(s_2)), c(\text{Izquierda}) + (0.8V(s_1) + 0.2V(s_2)) = \min\{1, 1\} = 1$

Iteración 2: $V(s_1) = 1 + (0.8 \times 1 + 0.2 \times 1) = 2$
$V(s_2) = \min\{1 + (0.8 \times 0 + 0.2 \times 1), 1 + (0.8 \times 1 + 0.2 \times 1)\} = \min\{1.2, 2\} = 1.2$

Iteración 3: $V(s_1) = 1 + (0.8 \times 1.2 + 0.2 \times 2) = 2.36$
$V(s_2) = \min\{1 + (0.8 \times 0 + 0.2 \times 1.2), 1 + (0.8 \times 2 + 0.2 \times 1.2)\} = \min\{1.24, 2.84\}$

Iteración 4: $V(s_1) = 1 + (0.8 \times 1.24 + 0.2 \times 2.36) = 2.464$
$V(s_2) = \min\{1 + (0.8 \times 0 + 0.2 \times 1.24), 1 + (0.8 \times 2.36 + 0.2 \times 1.24)\} = 1.248$

Ejemplo de Value Iteration

Iteración 1: $V(s_1) = c(\text{Derecha}) + (0.8V(s_2) + 0.2V(s_1)) = 1$
$V(s_2) = \min\{c(\text{Derecha}) + (0.8V(s_3) + 0.2V(s_2)), c(\text{Izquierda}) + (0.8V(s_1) + 0.2V(s_2)) = \min\{1, 1\} = 1$

Iteración 2: $V(s_1) = 1 + (0.8 \times 1 + 0.2 \times 1) = 2$
$V(s_2) = \min\{1 + (0.8 \times 0 + 0.2 \times 1), 1 + (0.8 \times 1 + 0.2 \times 1)\} = \min\{1.2, 2\} = 1.2$

Iteración 3: $V(s_1) = 1 + (0.8 \times 1.2 + 0.2 \times 2) = 2.36$
$V(s_2) = \min\{1 + (0.8 \times 0 + 0.2 \times 1.2), 1 + (0.8 \times 2 + 0.2 \times 1.2)\} = \min\{1.24, 2.84\}$

Iteración 4: $V(s_1) = 1 + (0.8 \times 1.24 + 0.2 \times 2.36) = 2.464$
$V(s_2) = \min\{1 + (0.8 \times 0 + 0.2 \times 1.24), 1 + (0.8 \times 2.36 + 0.2 \times 1.24)\} = 1.248$

Iteración 5: $V(s_1) = 1 + (0.8 \times 1.248 + 0.2 \times 2.464) = 2.4912$
$V(s_2) = \min\{1 + (0.8 \times 0 + 0.2 \times 1.248), 1 + (0.8 \times 2.464 + 0.2 \times 1.248)\} = 1.2496$

Ejemplo de Value Iteration

Iteración 1: $V(s_1) = c(\text{Derecha}) + (0.8V(s_2) + 0.2V(s_1)) = 1$
$V(s_2) = \min\{c(\text{Derecha}) + (0.8V(s_3) + 0.2V(s_2)), c(\text{Izquierda}) + (0.8V(s_1) + 0.2V(s_2)) =$
$\min\{1, 1\} = 1$

Iteración 2: $V(s_1) = 1 + (0.8 \times 1 + 0.2 \times 1) = 2$
$V(s_2) = \min\{1 + (0.8 \times 0 + 0.2 \times 1), 1 + (0.8 \times 1 + 0.2 \times 1)\} = \min\{1.2, 2\} = 1.2$

Iteración 3: $V(s_1) = 1 + (0.8 \times 1.2 + 0.2 \times 2) = 2.36$
$V(s_2) = \min\{1 + (0.8 \times 0 + 0.2 \times 1.2), 1 + (0.8 \times 2 + 0.2 \times 1.2)\} = \min\{1.24, 2.84\}$

Iteración 4: $V(s_1) = 1 + (0.8 \times 1.24 + 0.2 \times 2.36) = 2.464$
$V(s_2) = \min\{1 + (0.8 \times 0 + 0.2 \times 1.24), 1 + (0.8 \times 2.36 + 0.2 \times 1.24)\} = 1.248$

Iteración 5: $V(s_1) = 1 + (0.8 \times 1.248 + 0.2 \times 2.464) = 2.4912$
$V(s_2) = \min\{1 + (0.8 \times 0 + 0.2 \times 1.248), 1 + (0.8 \times 2.464 + 0.2 \times 1.248)\} = 1.2496$

Iteración 6: $V(s_1) = 1 + (0.8 \times 1.2496 + 0.2 \times 2.4912) = 2.49792$
$V(s_2) = \min\{1 + (0.8 \times 0 + 0.2 \times 1.2496), 1 + (0.8 \times 2.4912 + 0.2 \times 1.2496) = 1.24992$

Ejemplo de Value Iteration

Iteración 1: $V(s_1) = c(\text{Derecha}) + (0.8V(s_2) + 0.2V(s_1)) = 1$
$V(s_2) = \min\{c(\text{Derecha}) + (0.8V(s_3) + 0.2V(s_2)), c(\text{Izquierda}) + (0.8V(s_1) + 0.2V(s_2)) = \min\{1, 1\} = 1$

Iteración 2: $V(s_1) = 1 + (0.8 \times 1 + 0.2 \times 1) = 2$
$V(s_2) = \min\{1 + (0.8 \times 0 + 0.2 \times 1), 1 + (0.8 \times 1 + 0.2 \times 1)\} = \min\{1.2, 2\} = 1.2$

Iteración 3: $V(s_1) = 1 + (0.8 \times 1.2 + 0.2 \times 2) = 2.36$
$V(s_2) = \min\{1 + (0.8 \times 0 + 0.2 \times 1.2), 1 + (0.8 \times 2 + 0.2 \times 1.2)\} = \min\{1.24, 2.84\}$

Iteración 4: $V(s_1) = 1 + (0.8 \times 1.24 + 0.2 \times 2.36) = 2.464$
$V(s_2) = \min\{1 + (0.8 \times 0 + 0.2 \times 1.24), 1 + (0.8 \times 2.36 + 0.2 \times 1.24)\} = 1.248$

Iteración 5: $V(s_1) = 1 + (0.8 \times 1.248 + 0.2 \times 2.464) = 2.4912$
$V(s_2) = \min\{1 + (0.8 \times 0 + 0.2 \times 1.248), 1 + (0.8 \times 2.464 + 0.2 \times 1.248)\} = 1.2496$

Iteración 6: $V(s_1) = 1 + (0.8 \times 1.2496 + 0.2 \times 2.4912) = 2.49792$
$V(s_2) = \min\{1 + (0.8 \times 0 + 0.2 \times 1.2496), 1 + (0.8 \times 2.4912 + 0.2 \times 1.2496) = 1.24992$

Política óptima π^*: $\pi^*(s1) =$derecha, $\pi^*(s2) =$derecha

En este tema

Incertidumbre

- Razonamiento Probabilístico
- Redes Bayesianas
- Razonamiento Probabilístico en el tiempo
  - Introducción
  - Modelos de Markov
  - Modelos de Markov Ocultos (HMMs)
  - Procesos de Decisión de Markov (MDPs)
  - Procesos de Decisión de Markov Parcialmente Observables (POMDPs)
- Lógica Borrosa

Procesos de Decisión de Markov Parcialmente Observables (POMDPs)

- Proceso de Markov + observabilidad parcial = HMM
- Proceso de Markov + acciones = MDP
- Proceso de Markov + observabilidad parcial + acciones = HMM + acciones = MDP + observabilidad parcial = POMDP

acciones	observabilidad total	observabilidad parcial
no	Proceso de Markov	HMM
si	MDP	POMDP

Créditos

- Material UC3M de cursos anteriores
- Libro *Artificial Intelligence: a modern approach*. Russell and Norving. Prentice Hall. 2003
- Material docente de Andrew M. Moore. Carnegie Mellon University. `http://www.cs.cmu.edu/~awm/tutorials`
- Video de Sebastian Thrun. Stanford University. `https://www.youtube.com/watch?v=DgH6NaJHfVQ&list=PLTnAYbE3sS--_h8N7dTp9vm8MRT2m1IOD`

En este tema

Incertidumbre

Razonamiento Probabilístico
Redes Bayesianas
Razonamiento Probabilístico en el tiempo
- Introducción
- Modelos de Markov
- Modelos de Markov Ocultos (HMMs)
- Procesos de Decisión de Markov (MDPs)
- Procesos de Decisión de Markov Parcialmente Observables (POMDPs)

Lógica Borrosa