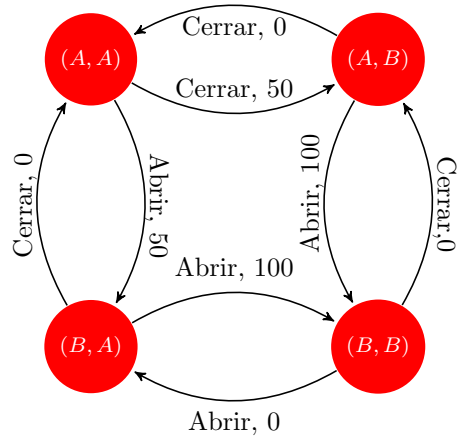


## Ejemplo de Aprendizaje por refuerzo

16 de diciembre de 2008

Queremos hacer que un sistema de control aprenda a controlar una válvula que gobierna la presión y la temperatura de un sistema. Consideraremos que la presión y la temperatura pueden tomar los valores **alta** y **baja**. De esta manera tenemos cuatro estados definidos por las combinaciones de los valores de estas dos variables. El sistema de control puede hacer dos acciones: Abrir, Cerrar.

La función  $\delta$  de cambio de estado y la función  $r$  de refuerzo son las del siguiente grafo:



Supondremos que para entrenar el sistema aprenderemos a partir de tres secuencias de entrenamiento:

1. Estado Inicial=(A,A), acciones={cerrar, cerrar, abrir, abrir}
2. Estado Inicial=(A,B), acciones={cerrar, cerrar, abrir}
3. Estado Inicial=(B,B), acciones={cerrar, abrir}

El valor del parámetro  $\gamma$  es de 0.9.

## Solución

La tabla que representa la función  $\hat{Q}$  inicial es la siguiente:

$\hat{Q}$	Acción	
	Abrir	Cerrar
(A,A)	0	0
(A,B)	0	0
(B,A)	0	0
(B,B)	0	0

### 1. Primer entrenamiento

a) Estado=(A,A), acción=Cerrar

$$\begin{aligned}\hat{Q}((A, A), cerrar) &= r((A, A), cerrar) + \gamma \max_{ac \in \{abrir, cerrar\}} \hat{Q}((A, B), ac) \\ &= 50 + 0.9 \cdot \max\{0, 0\} \\ &= 50\end{aligned}$$

$\hat{Q}$	Acción	
	Abrir	Cerrar
(A,A)	0	50
(A,B)	0	0
(B,A)	0	0
(B,B)	0	0

b) Estado=(A,B), acción=Cerrar

$$\begin{aligned}\hat{Q}((A, B), cerrar) &= r((A, B), cerrar) + \gamma \max_{ac \in \{abrir, cerrar\}} \hat{Q}((A, A), ac) \\ &= 0 + 0.9 \cdot \max\{0, 50\} \\ &= 45\end{aligned}$$

$\hat{Q}$	Acción	
	Abrir	Cerrar
(A,A)	0	50
(A,B)	0	45
(B,A)	0	0
(B,B)	0	0

c) Estado=(A,A), acción=Abrir

$$\begin{aligned}\hat{Q}((A, A), abrir) &= r((A, A), abrir) + \gamma \max_{ac \in \{abrir, cerrar\}} \hat{Q}((B, A), ac) \\ &= 50 + 0.9 \cdot \max\{0, 0\} \\ &= 50\end{aligned}$$

$\hat{Q}$	Acción	
	Abrir	Cerrar
(A,A)	50	50
(A,B)	0	45
(B,A)	0	0
(B,B)	0	0

d) Estado=(B,A), acción=Abrir

$$\begin{aligned}\hat{Q}((B, A), abrir) &= r((B, A), abrir) + \gamma \max_{ac \in \{abrir, cerrar\}} \hat{Q}((B, B), ac) \\ &= 100 + 0.9 \cdot \max\{0, 0\} \\ &= 100\end{aligned}$$

$\hat{Q}$	Acción	
	Abrir	Cerrar
(A,A)	50	50
(A,B)	0	45
(B,A)	100	0
(B,B)	0	0

## 2. Segundo entrenamiento

a) Estado=(A,B), acción=Cerrar

$$\begin{aligned}\hat{Q}((A, B), cerrar) &= r((A, B), cerrar) + \gamma \max_{ac \in \{abrir, cerrar\}} \hat{Q}((A, A), ac) \\ &= 0 + 0.9 \cdot \max\{50, 50\} \\ &= 45\end{aligned}$$

$\hat{Q}$	Acción	
	Abrir	Cerrar
(A,A)	50	50
(A,B)	0	45
(B,A)	100	0
(B,B)	0	0

b) Estado=(A,A), acción=Cerrar

$$\begin{aligned}\hat{Q}((A, A), cerrar) &= r((A, A), cerrar) + \gamma \max_{ac \in \{abrir, cerrar\}} \hat{Q}((A, B), ac) \\ &= 50 + 0.9 \cdot \max\{0, 45\} \\ &= 90.5\end{aligned}$$

$\hat{Q}$	Acción	
	Abrir	Cerrar
(A,A)	50	90.5
(A,B)	0	45
(B,A)	100	0
(B,B)	0	0

c) Estado=(A,B), acción=Abrir

$$\begin{aligned}\hat{Q}((A, B), abrir) &= r((A, B), abrir) + \gamma \max_{ac \in \{abrir, cerrar\}} \hat{Q}((B, B), ac) \\ &= 100 + 0.9 \cdot \max\{0, 0\} \\ &= 100\end{aligned}$$

$\hat{Q}$	Acción	
Estado	Abrir	Cerrar
(A,A)	50	90.5
(A,B)	100	45
(B,A)	100	0
(B,B)	0	0

### 3. Tercer entrenamiento

a) Estado=(B,B), acción=Cerrar

$$\begin{aligned} \hat{Q}((B, B), cerrar) &= r((B, B), cerrar) + \gamma \max_{ac \in \{abrir, cerrar\}} \hat{Q}((A, B), ac) \\ &= 0 + 0.9 \cdot \max\{100, 45\} \\ &= 90 \end{aligned}$$

$\hat{Q}$	Acción	
Estado	Abrir	Cerrar
(A,A)	50	90.5
(A,B)	100	45
(B,A)	100	0
(B,B)	0	90

b) Estado=(A,B), acción=Abrir

$$\begin{aligned} \hat{Q}((A, B), abrir) &= r((A, B), abrir) + \gamma \max_{ac \in \{abrir, cerrar\}} \hat{Q}((B, B), ac) \\ &= 100 + 0.9 \cdot \max\{0, 90\} \\ &= 181 \end{aligned}$$

$\hat{Q}$	Acción	
Estado	Abrir	Cerrar
(A,A)	50	90.5
(A,B)	181	45
(B,A)	100	0
(B,B)	0	90

La función  $\hat{Q}$  aprendida sería:

