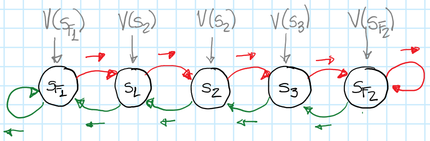
Ecuaciones de Bellman Deterministas: Ejemplo

sábado, 19 de marzo de 2022 10:39 a. m.



sF1	s1	s2	s3	sF2
-10	0	-0.4	-0.4	10

	sF1	s1	s2	s3	sF2
$F_\pi(S) =$		→	→	→	

$\gamma = 0.9$

M_T Determinista

$$V(s) = \overline{f_{RA}(\tau)}\Big|_{s(0)=s_0}$$
$$= f_R(s,a,s_f) + \gamma V(s_f)$$
$$= f_R\left(s,a,f_{M_T}(s,a)\right) + \gamma V\left(f_{M_T}(s,a)\right)$$

con: $a = f_\pi(s)$

$$V(s) = f_R\left(s, f_\pi(s), f_{M_T}(s,a)\right) + \gamma V\left(f_{M_T}(s,a)\right)$$

$V(s_1) = -0.4 + \gamma V(s_2)$

$V(s_2) = -0.4 + \gamma V(s_3)$

$V(s_3) = 10 + \gamma V(s_{F_2})$

$V(s_{F_1}) = -10 + \gamma V(s_{F_1})$

$V(s_{F_2}) = 10 + \gamma V(s_{F_2})$

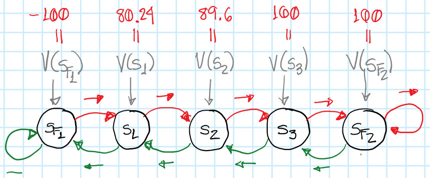
$V(s_1) = -0.4 + \gamma V(s_2)$
$= -0.4 + (0.9)(89.6)$
$= 80.24$

$V(s_3) = 10 + \gamma V(s_{F_2})$
$= 10 + (0.9)(100)$
$= 10 + 90$
$= 100$

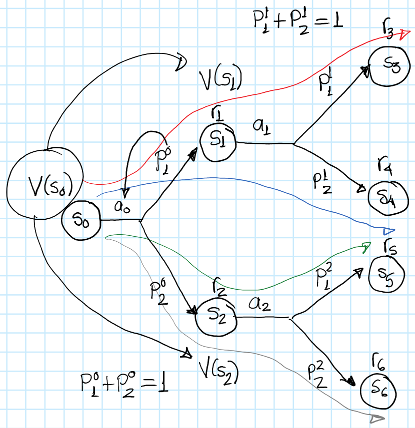
$V(s_{F_2}) = 10 + \gamma V(s_{F_2})$
$V(s_{F_2}) - \gamma V(s_{F_2}) = 10$
$[1-\gamma]\, V(s_{F_2}) = 10$
$V(s_{F_2}) = \frac{10}{[1-\gamma]} = \frac{10}{[1-0.9]} = \frac{10}{0.1} = 100$

$V(s_{F_1}) = -10 + \gamma V(s_{F_1})$
$V(s_{F_1}) - \gamma V(s_{F_1}) = -10$
$[1-\gamma]\, V(s_{F_1}) = -10$
$V(s_{F_1}) = \frac{-10}{[1-\gamma]} = -100$

$V(s_2) = -0.4 + \gamma V(s_3)$
$= -0.4 + (0.9)(100)$
$= -0.4 + 90$
$= 89.6$



Obtención de las Ecuaciones de Bellman NO deterministas



$P_1^0 P_1^1 [r_1 + \gamma r_3]$

$P_1^0 P_2^1 [r_1 + \gamma r_4]$

$P_2^0 P_1^2 [r_2 + \gamma r_5]$

$P_2^0 P_2^2 [r_2 + \gamma r_6]$

$s_f \sim P_{M_T}(s_f \mid s, a)$

$P_1^0 = P_{M_T}(s_1 \mid s_0, a_0)$

$P_2^0 = P_{M_T}(s_2 \mid s_0, a_0)$

$$V(s_0) = P_1^0 P_1^1 [r_1 + \gamma r_3] + P_1^0 P_2^1 [r_1 + \gamma r_4] + P_2^0 P_1^2 [r_2 + \gamma r_5] + P_2^0 P_2^2 [r_2 + \gamma r_6]$$

$$= P_1^0 \left[P_1^1 [r_1 + \gamma r_3] + P_2^1 [r_1 + \gamma r_4] \right] + P_2^0 \left[P_1^2 [r_2 + \gamma r_5] + P_2^2 [r_2 + \gamma r_6] \right]$$

$$= P_1^0 \left[(\underbrace{P_1^1 + P_2^1}_{1}) r_1 + \gamma (P_1^1 r_3 + P_2^1 r_4) \right] + P_2^0 \left[(\underbrace{P_1^2 + P_2^2}_{1}) r_2 + \gamma (P_1^2 r_5 + P_2^2 r_6) \right]$$

$$= P_1^0 r_1 + \gamma P_1^0 \underbrace{(P_1^1 r_3 + P_2^1 r_4)}_{V(s_1)} + P_2^0 r_2 + \gamma P_2^0 \underbrace{(P_1^2 r_5 + P_2^2 r_6)}_{V(s_2)}$$

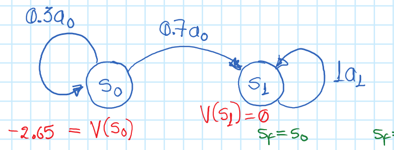
$$= P_1^0 r_1 + \gamma P_1^0 V(s_1) + P_2^0 r_2 + \gamma P_2^0 V(s_2)$$

$$= P_1^0 [r_1 + \gamma V(s_1)] + P_2^0 [r_2 + \gamma V(s_2)]$$

$$= P_{M_T}(s_1 \mid s_0, a_0) [r_1 + \gamma V(s_1)] + P_{M_T}(s_2 \mid s_0, a_0) [r_2 + \gamma V(s_2)]$$

$$V(s_0) = \sum_{i=1}^{2} P_{M_T}(s_i \mid s_0, a_0) [r_i + \gamma V(s_i)]$$

Ecuaciones de Bellman NO deterministas: Ejemplo



$P_{M_T}(s_f|s,a) =$

$s_f=s_0$	a_0	a_1
s_0	0.3	0
s_1	0	0

$s_f=s_1$	a_0	a_1
s_0	0.7	0
s_1	0	1

$f_R(s,a,s_f) =$

$s_f=s_0$	a_0	a_1
s_0	0	1
s_1	1	2

$s_f=s_1$	a_0	a_1
s_0	-3	-1
s_1	1	0

$f_\pi(s_f) =$

s_0	a_0
s_1	a_1

$\gamma = 0.7$

M_T NO determinista

$$V(s) = \overline{f_{RA}(\tau)}\Big|_{s(0)=s}$$
$$= \overline{f_R(s,a,s_f) + \gamma V(s_f)}$$
$$= \sum_{s_f\in S} P_{M_T}(s_f|s,a)[f_R(s,a,s_f) + \gamma V(s_f)]$$

con: $a = f_\pi(s)$

$$V(s) = \sum_{s_f\in S} P_{M_T}(s_f|s,f_\pi(s))\,[f_R(s,f_\pi(s),s_f) + \gamma V(s_f)]$$

$$V(s_0) = \sum_{i=0}^{1} P_{M_T}(s_i|s_0,a_0)\left[f_R(s_0,a_0,s_i) + \gamma V(s_i)\right]$$

$$= \underbrace{P_{M_T}(s_0|s_0,a_0)}_{0.3}\Big[\underbrace{f_R(s_0,a_0,s_0)}_{0} + \gamma V(s_0)\Big] + \underbrace{P_{M_T}(s_1|s_0,a_0)}_{0.7}\Big[\underbrace{f_R(s_0,a_0,s_1)}_{-3} + \gamma V(s_1)\Big]$$

$$= (0.3)\left[0 + (0.7)V(s_0)\right] + (0.7)\left[(-3) + \gamma V(s_1)\right]$$

$$V(s_0) = 0.21\,V(s_0) - 2.1 + 0.49\,V(s_1)$$

$$V(s_0) - 0.21\,V(s_0) = -2.1 + 0.49\,V(s_1)$$

$$\boxed{0.79\,V(s_0) = -2.1 + 0.49\,V(s_1)}$$

$$V(s_1) = \sum_{i=0}^{1} P_{M_T}(s_i|s_1,a_1)\left[f_R(s_1,a_1,s_i) + \gamma V(s_i)\right]$$

$$= \underbrace{P_{M_T}(s_0|s_1,a_1)}_{0}\Big[f_R(s_1,a_1,s_0) + \gamma V(s_0)\Big]^{0} + \underbrace{P_{M_T}(s_1|s_1,a_1)}_{1}\Big[\underbrace{f_R(s_1,a_1,s_1)}_{0} + \gamma V(s_1)\Big]$$

$$V(s_1) = 1\left[(0.7)V(s_1)\right]$$

$$V(s_1) - 0.7\,V(s_1) = 0$$

$$\boxed{0.3\,V(s_1) = 0}$$

$$V(s_1) = \frac{0}{0.3} = 0$$

$$0.79\,V(s_0) = -2.1 + \cancel{0.49(0)}^{\,0}$$

$$V(s_0) = -\frac{2.1}{0.79} = -2.65$$