MDP

① Bellman-EQ

* $U(s) = \max_{a \in A(s)} \sum_{s'} P(s'|s,a)\left[R(s,a,s') + \gamma U(s')\right]$

$= \max_{a \in A(s)} R(s,a) + \sum_{s'} P(s'|s,a)\, \gamma U(s')$

$= R(s) + \max_{a \in A(s)} \sum_{s'} P(s'|s,a)\, \gamma U(s')$

* $\pi(s)$ = same, but use $\boxed{\arg\max_{a \in A(s)}}$ instead of $\boxed{\max_{a \in A(s)}}$

② Geometric sum

↳ $\sum_{i=1}^{n} \gamma^i r = \frac{r(1-\gamma^n)}{1-\gamma}$, $\sum_{i=0}^{\infty} \gamma^i r \leq \frac{R_{max}}{1-\gamma}$

---

* **Value iteration** →

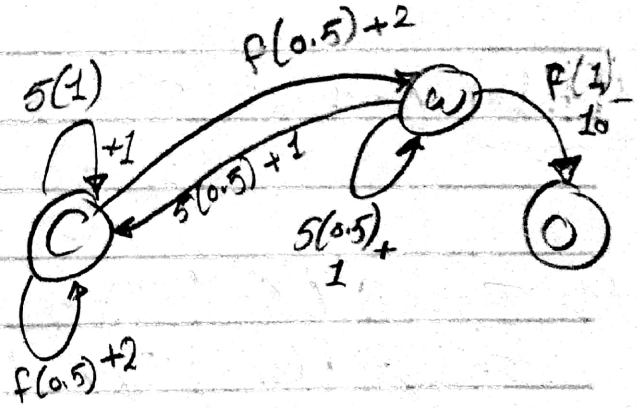
* حدد القانون حسب شكل الـ R بتاعتك
* حط قيمة initial لـ U، خليها بـ Zero لو مدا كش حاجة.
* أكمل عدد iterations معين أو لحد فرق أقل من threshold

* **Policy iteration** →

* حدد القانون حسب شكل الـ R بتاعتك
* حط Initial Actions سواء من عندك أو Given
* احسب الـ U() بس المرادي مفيش Max، هتحسب على الـ action اللي عملته.
* هيطلع عندك كذا معادلات في مجاهيل حلهم مع بعض عشان تطلع الـ U.
* استعمل معادلة الـ π() عشان تشوف الـ argmax وده الـ action الجديد.
* كمل كده لحد ما تـ converge

Sh7

17.4 $\gamma = 0.5$



$$U(S) = \max_{a \in A(S)} R(S,A) + \sum P(S'/S,a)\,\gamma U(S')$$

※ Value Iteration:

l	U(C)	U(W)	U(O)
0	0	0	0
1	Max [1 + P(C/C,S) * γU(C), 2 + [P(C/C,F) * γU(C) + P(W/C,F) * γU(W)]] = 2	Max [1 + [P(W/W,S) * γU(W) + P(C/W,S) * γU(C)], -10 [P(O/W,F) * γU(O)]] = 1	No Actions = 0
2	Max [1 + 1 * γU(C), 2 + [0.5 * γU(C) + 0.5 * γU(W)]] = 2.75	Max [1 + [0.5 * γU(W) + 0.5 * γU(C)], -10 [1 * γU(O)]] = 1.75	No Actions = 0

※ Policy Extraction [No iterations]

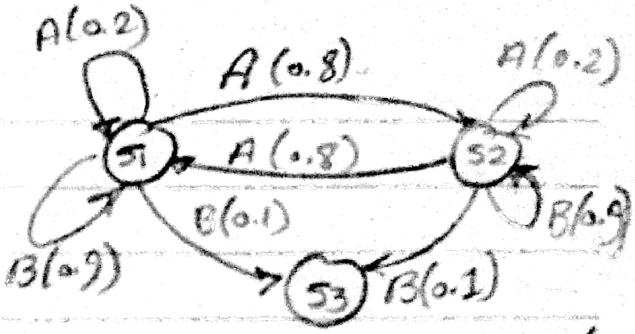
$$\pi(S) = \arg\max_{a \in A(S)} R(S,A) + \sum P(S'/S,a)\,\gamma U(S')$$

$\pi(C) = \arg\max$ [1 + 1 * γU(C), 2 + [0.5 * γU(C) + 0.5 * γU(W)]] (2.75, 1.75) = 1

$\pi(W) = \arg\max$ [1 + [0.5 * γU(W) + 0.5 * γU(C)], -10 [1 * γU(O)]] = 5

$\pi(O) = -$

17.1. $\gamma = 1$, $R(S_1) = -1$, $R(S_2) = -2$
$R(S_3) = 0$ ← Terminal



(1) $S_1 \rightarrow B$ [Go to S_3 or Stay in S_1], [A] is bad bec more Penalty
$S_2 \rightarrow A$ [Go to S_1 and stuck there is better], [B] why bad?
Bec it has lower Prob to succeed!

(2) Value Iteration $[U(S) = R(S) + \max_{a \in A(S)} [P(S'|S,a)\,\gamma\,U(S)]$

ℓ	$U(S_1)$	$U(S_2)$	$U(S_3)$
0	0	0	0
1	$-1 + \max\begin{bmatrix} 0.2*\gamma U(S_1) + 0.8*\gamma U(S_2), \\ 0.9*\gamma U(S_1) + 0.1*\gamma U(S_3) \end{bmatrix}$	$-2 + \max\begin{bmatrix} 0.2*\gamma U(S_2) + 0.8*\gamma U(S_1), \\ 0.9*\gamma U(S_2) + 0.1*\gamma U(S_3) \end{bmatrix}$	No Actions
	$= -1$	$= -2$	$= 0$
2	$-1 + \max\begin{bmatrix} 0.2*-1 + 0.8*-2, \\ 0.9*-1 + 0.1*0 \end{bmatrix}$	$-2 + \max\begin{bmatrix} 0.2*-2 + 0.8*-1, \\ 0.9*-2 + 0.1*0 \end{bmatrix}$	
	$= -1.9$	$= -3.2$	$= 0$
3	$= -2.71$	$= -4.16$	$= 0$

* Policy Extraction

$\pi(S_1) = \arg\max \begin{bmatrix} 0.2*U(S_1) + 0.8\,U(S_2), \\ 0.9*U(S_1) + 0.1\,U(S_3) \end{bmatrix} = B$

$\pi(S_2) = \arg\max \begin{bmatrix} 0.2*U(S_2) + 0.8\,U(S_1), \\ 0.9*U(S_2) + 0.1\,U(S_3) \end{bmatrix} = A$

③ Policy Iteration

ℓ		S_1	S_2	S_3
	U	—	—	—
0	π	B	B	—
1	U	$-1+(0.1\times\gamma U(S_3)+0.9\times\gamma U(S_1))$	$-2+(0.1\times\gamma U(S_3)+0.9\gamma U(S_2))$	—
		$=-10$	$=-20$	
	π	$\text{argmax}\begin{bmatrix}0.8U(S_2)+0.2U(S_1),\\ 0.1U(S_3)+0.9U(S_2)\end{bmatrix}$	$\text{argmax}\begin{bmatrix}0.8U_{S_1}+0.2U_{S_2},\\ 0.1U_{S_3}+0.9U_{S_2}\end{bmatrix}$	—
		$=B$	$=A$	
2	U	$-1+(0.1U_{S_3}+0.9U_{S_1})$	$-2+(0.8U_{S_1}+0.2U_{S_2})$	
		$=-10$	$=-12.5$	
	π	$\text{argmax}\begin{bmatrix}0.8U_{S_2}+0.2U_{S_1},\\ 0.1U_{S_3}+0.9U_{S_1}\end{bmatrix}$	$\text{argmax}\begin{bmatrix}0.8U_{S_1}+0.2U_{S_2},\\ 0.1U_{S_3}+0.9U_{S_2}\end{bmatrix}$	
		$=B$ → Converged	$=A$ → Converged	

④ What if the initial action was 'A' for both?

$$\begin{cases} U(S_1) = -1+0.8(\gamma U_{S_2})+0.2(\gamma U_{S_1}) \\ U(S_2) = -2+0.8(\gamma U_{S_1})+0.2(\gamma U_{S_2}) \end{cases} \quad \begin{matrix} U(S_1)\to x \\ U(S_2)\to y \end{matrix}$$

$$\rightarrow \begin{cases} -0.8x+0.8y=1 \\ 0.8x-0.8y=2 \end{cases}$$ Two // lines, No Sol!

$\gamma\neq 0$

$$\begin{matrix}(0.2\gamma-1)x+0.8\gamma y=1 \\ 0.8\gamma x+(0.2\gamma-1)y=2\end{matrix} \longrightarrow \begin{pmatrix}0.2\gamma-1 & 0.8\\ 0.8 & 0.2\gamma-1\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}1\\2\end{pmatrix}$$

* Have a Sol iff $\nabla\neq 0 \longrightarrow \begin{matrix}(0.2\gamma-1)^2\\ -(0.8)^2\end{matrix}\neq 0 \longrightarrow \gamma\neq 1$ $\gamma\neq 1$ → But, $0<\gamma<1$ So, No Sol $\forall\gamma$