Lecture - 9

* Policy evaluation

- Given a policy π, find V^π
- P & R are known

Method:

① Solve the system of ~~linear~~ eqn produced by the Bellman equation

$$V^\pi(s) = \sum_a \pi(s,a) \sum_{s'} P(s,a,s') \left[R(s,a,s') + \gamma V^\pi(s') \right]$$

② Dynamic Programming

- Sequence of approximations of V^π

e.g. $V_0^\pi, V_1^\pi, V_2^\pi, \ldots$

V_0^π → Chosen arbitrarily except $V_0^\pi(s) = 0$ i.e. for terminal state

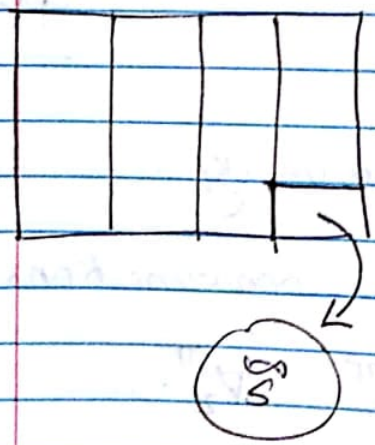
Dynamic programming

③
$$\hat{V}^{\pi}_{k+1}(s) = \sum_{a} \pi(s,a) \sum_{s'} P(s,a,s') \left[R(s,a,s') + \gamma \hat{V}^{\pi}_{k}(s') \right]$$

Properties –

- $\hat{V}^{\pi}_{k} = V^{\pi}$ is a fixed-point.
- $\hat{V}^{\pi}_{k} \to V^{\pi}$ as $k \to \infty$ for finite MDPs
- One pass over state space is a "full backup". A single state update is called "backup"

Example :–



$R_t = -1$ always

$\gamma = 1$

$\pi =$

↓	↓	↓	↓
↓	↓	↓	↓
↓	↓	↓	↓
→	→	→	

$\hat{V}^{\pi}_{0} =$

0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0

$\Rightarrow \hat{V}^{\pi}_{1} =$

			s'
		s''	s'''
		0	s''''

$$V_1^{\pi}(s) = s'\left(-1 + \gamma \hat{V}_k^{\pi}(s)\right) \quad \rightarrow \text{Prob.} = \text{zero}$$

$$+ \ldots$$

$$+ \underbrace{\left(-1 + \gamma \hat{V}_k^{\pi}(s''')\right)}_{0} \quad \rightarrow \text{Prob} = 1$$

$$\therefore \hat{V}_1^{\pi} = \begin{bmatrix} -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & 0 \end{bmatrix}$$

$$\text{Now, } \hat{V}_2^{\pi} = \begin{bmatrix} -2 & -2 & -2 & -2 \\ -2 & -2 & -2 & -2 \\ -2 & -2 & -2 & -1 \\ -2 & -2 & -1 & 0 \end{bmatrix}$$

$$\ldots \hat{V}_6^{\pi} = \begin{bmatrix} -6 & -5 & -4 & -3 \\ -5 & -4 & -3 & -2 \\ -4 & -3 & -2 & -1 \\ -3 & -2 & -1 & 0 \end{bmatrix}$$

* In-place implementation

- when updating $\hat{V}_k^{\pi}(s)$ store it back in same table
- update states in any order
- Asynchronous updates

* Policy improvement

$$\hat{q}^{\pi}_{k+1}(s,a) = \sum P(s,a,s')\left[R(s,a,s') + \gamma \pi(s',a') \cdot \hat{q}^{\pi}_{k}(s',a'\right.$$

Let π & π' be deterministic polici such that for all states

$$q^{\pi}(s,\pi'(s)) \geq v^{\pi}(s) = q^{\pi}(s,\pi(s))$$

Then $\pi' \geq \pi$

$$\forall s \in S \quad , \quad v^{\pi'}(s) \geq v^{\pi}(s)$$