Lec-8, IT567, 24-25

$$v_\pi(s) = E_\pi[R_{t+1} + \gamma G_{t+1} \mid S_t = s]$$

$$T_1 = E_\pi[R_{t+1} \mid S_t = s] \quad ; \quad T_2 = \gamma E_\pi[G_{t+1} \mid S_t = s]$$

$$v_\pi(s) = T_1 + T_2 .$$

$$T_1 = \sum_r r\, p(r|s) = \sum_{s'} p(s'|s) \sum_r r\, p(r|s,s') \text{ — May not be reqd.}$$

$$= \sum_{s'} \sum_r r\, \underbrace{p(s',r|s)}_{\Downarrow \; \sum_a \pi(a|s)\, p(s',r|s,a)} = \sum_{s'} \sum_r \sum_a r\, \pi(a|s)\, p(s',r|s,a)$$

$$T_2 = \gamma \underbrace{\sum_{s'} p(s'|s)}_{\Downarrow} v_\pi(s') \text{ — from last slide of Lec 7.}$$

$$\sum_{s'} \sum_a \sum_r \pi(a|s)\, p(s',r|s,a) \times \gamma v_\pi(s')$$

Let take $S = \{s_1, s_2, s_3, s_4\}$

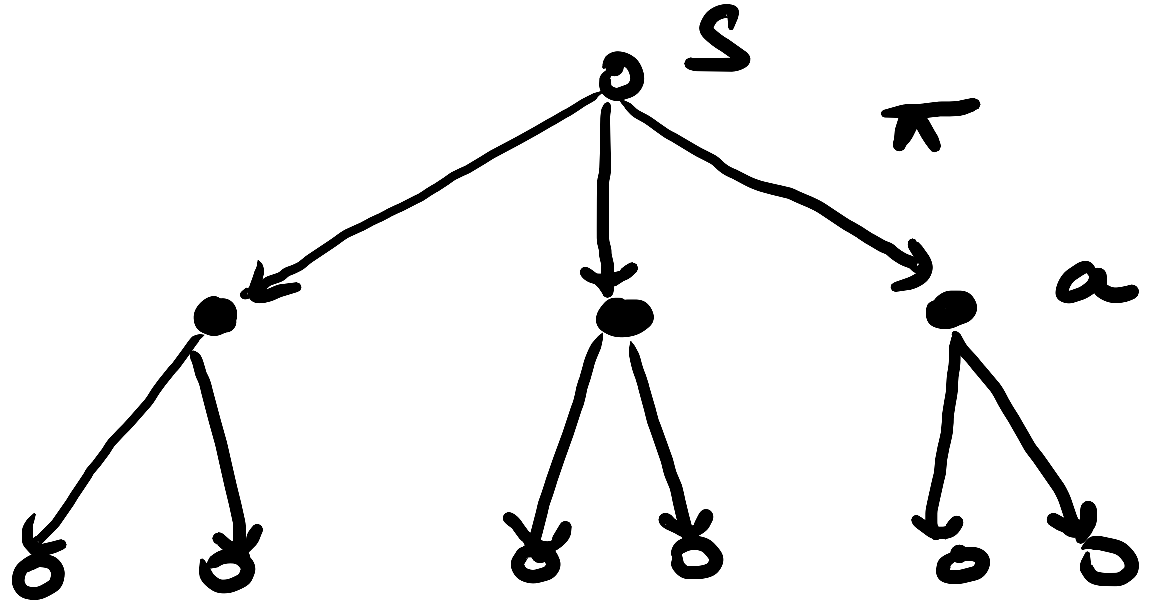
$V_\pi(s_1) \rightarrow V_\pi(s_2), V_\pi(s_3), V_\pi(s_4), V_\pi(s_1)$

$V_\pi(s_2) \rightarrow V_\pi(s_1), V_\pi(s_3), V_\pi(s_4), V_\pi(s_2)$

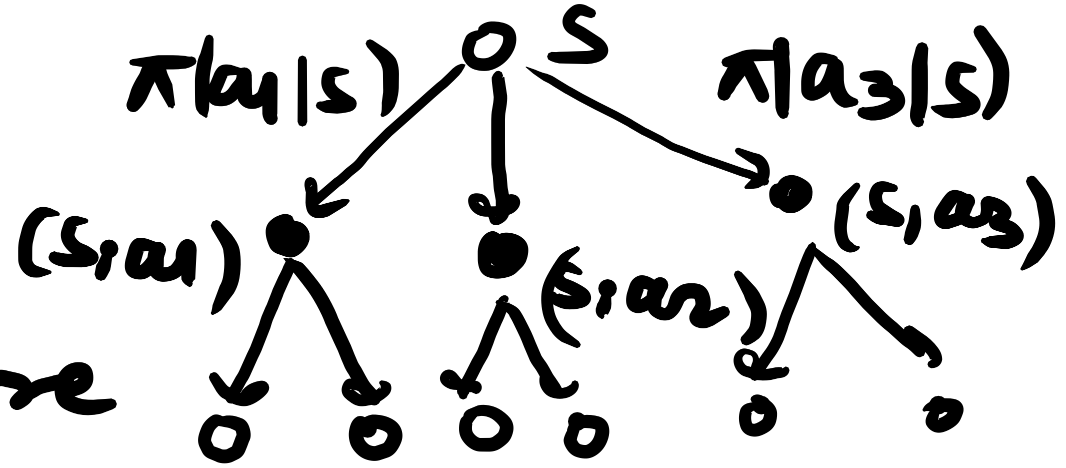
$\vdots$

because the environment can take the agent from the current state to the same state. (self-loop)

Backup diagram for V_π



in this example, let the no. of possible actions in s be 3, a_1, a_2 & a_3.



after taking an action, suppose you can go to 2 possible states

$v_\pi(s) = E_\pi[G_t \mid S_t = s]$

$q_\pi(s,a) = E_\pi[G_t \mid S_t = s, A_t = a]$

1. find $v_\pi(s)$ in terms of $q_\pi(s,a)$. → $\sum_a \pi(a|s)\, q_\pi(s,a)$
2. find $q_\pi(s,a)$ " " $v_\pi(s)$ & $p(s', r|s,a)$. → H.W.

In the backup diagram, each open circle represents a state, while each solid circle rep. a state-action pair.

Bellman equality :- states that the value of the state you are currently in = (discounted) value of the expected next state plus the reward expected along the way.

The state-value function $V_\pi(s)$ is the 'unique solution' to its Bellman equation (BE).

Optimal policies & optimal value functions.

Aim (roughly) is to find a policy that achieves a lot of reward over the long run.

Defn:- A policy π is defined to be better than or equal to a policy π', if its expected return is $\geq$ that of π'

$\forall$ states

$$\pi \geq \pi' \quad \text{iff} \quad V_\pi(s) \geq V_{\pi'}(s) \quad \forall s \in \mathcal{S}.$$