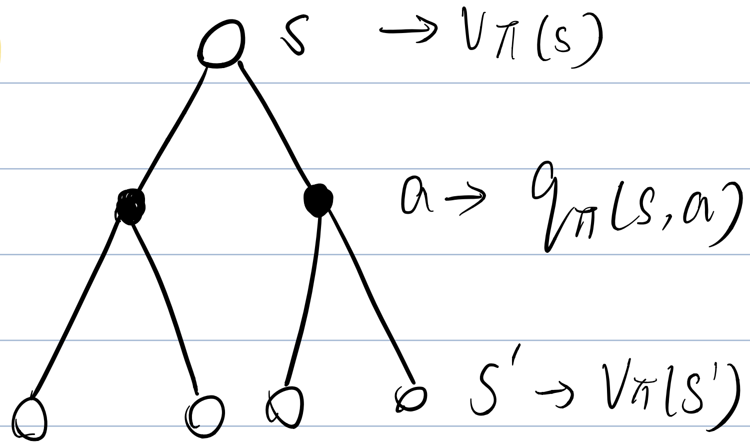
Markov Property

Markov Process → Markov reward process.
↓
Markov Decision Process.

MDP:



$M = \langle S, A, P, R, \gamma \rangle$ policy $\pi$

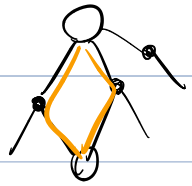
当只关注 State 时

$MP = \langle S, P^\pi \rangle$

当关于 State 与 Reward 时

$MRP = \langle S, P^\pi, R^\pi, \gamma \rangle$

$S \rightarrow S'$



$$P^\pi_{s,s'} = \sum_{a \in A} \pi(a|s) \cdot P^a_{ss'}$$

S.

$$R^\pi_s = \sum \pi(a|s) R^a_s$$

$R^a_s$ 或为

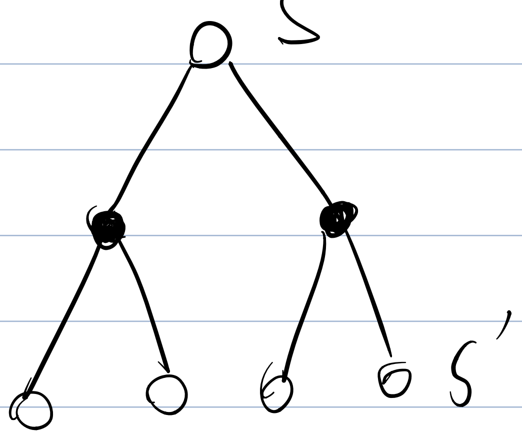
[illegible]

$a \in A$ 路径上的值)

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

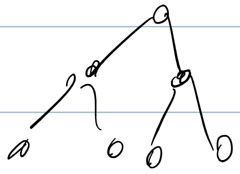
$$q_\pi(s,a) = E_\pi[G_t \mid S_t = s, A_t = a]$$

寻找与 $V_\pi(s')$ 关系、
与 $q_\pi(s',a)$ 的关系



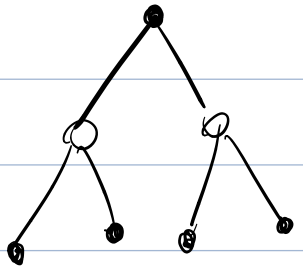
$$V_\pi(s) = E_\pi[R_{t+1} + \gamma V_\pi(S_{t+1}) \mid S_t = s]$$

$$V_\pi(s) = \sum_{a \in A} \pi(a|s) \cdot q_\pi(s,a)$$



$$q_\pi(s,a) = E_\pi(R_{t+1} + \gamma q_\pi(S_{t+1}, A_{t+1}) \mid S_t = s, A_t = a)$$

$$= R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a V_\pi(s')$$



$\Rightarrow$

$$V_\pi(s) = \sum_{a \in A} \pi(a|s) \cdot \left(R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a \cdot V_\pi(s')\right)$$

$$q_\pi(s,a) = R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a \cdot \sum_{a \in A} \pi(a|s) \cdot q_\pi(s,a)$$

寻找最优解= $\Leftrightarrow$ 最优 policy $\pi$

Intuition