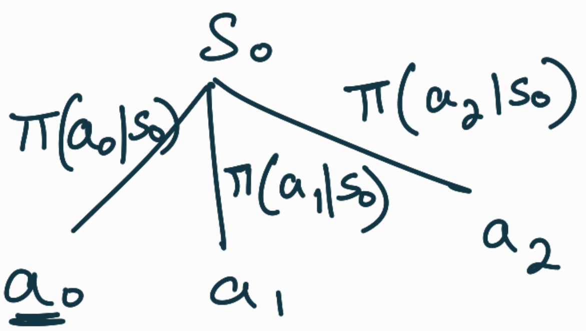
$$v_\pi(s) = E[G_t \mid S_t = s]$$

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

$$S = \{s_0, s_1\}$$

$$A = \{a_0, a_1, a_2\}$$

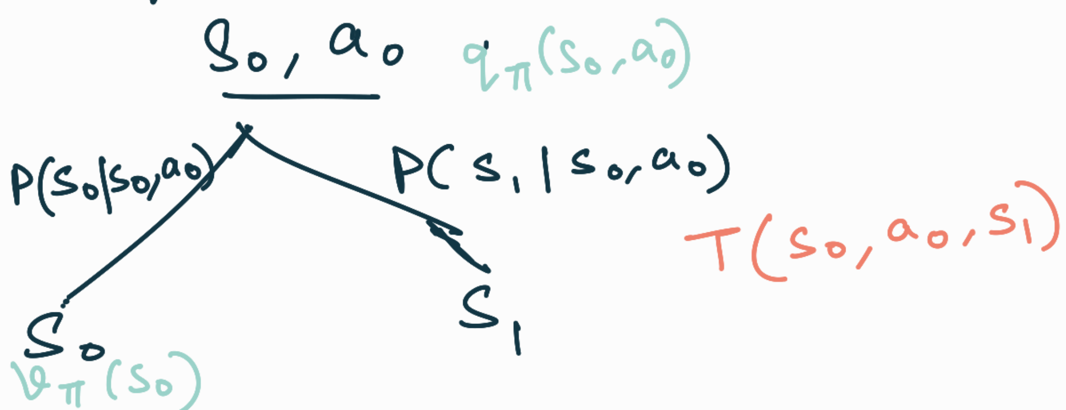
* one level back up diagram starting from state s_0.



$$v_\pi(s_0) = \pi(a_0|s_0)\, q_\pi(s_0,a_0) + \pi(a_1|s_0)\, q_\pi(s_0,a_1) + \pi(a_2|s_0)\, q_\pi(s_0,a_2)$$

$$= \sum_{a \in A} \pi(a|s_0)\, q_\pi(s_0,a) \rightarrow (1)$$

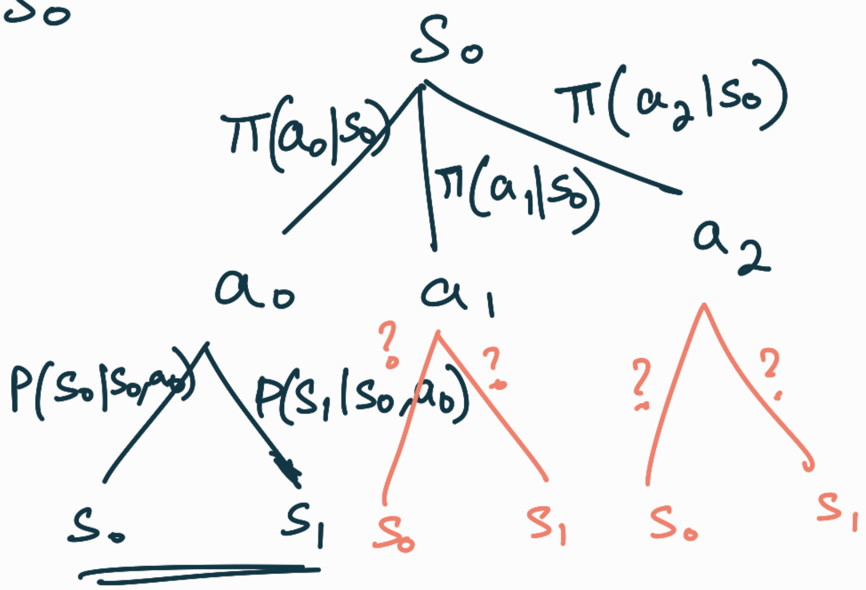
* Starting from state s_0 and taking action a_0.



$$q_\pi(s_0, a_0) = P(s_0|s_0,a_0)\left(r(s_0,a_0,s_0) + \gamma\, v_\pi(s_0)\right)$$
$$+ P(s_1|s_0,a_0)\left(r(s_1,a_0,s_0) + \gamma\, v_\pi(s_1)\right)$$
$$= \sum_{s' \in S} P(s'|s_0,a_0)\left[r(s_0,a_0,s') + \gamma\, v_\pi(s')\right]$$

2-Step backup diagram

* S_0



From ①

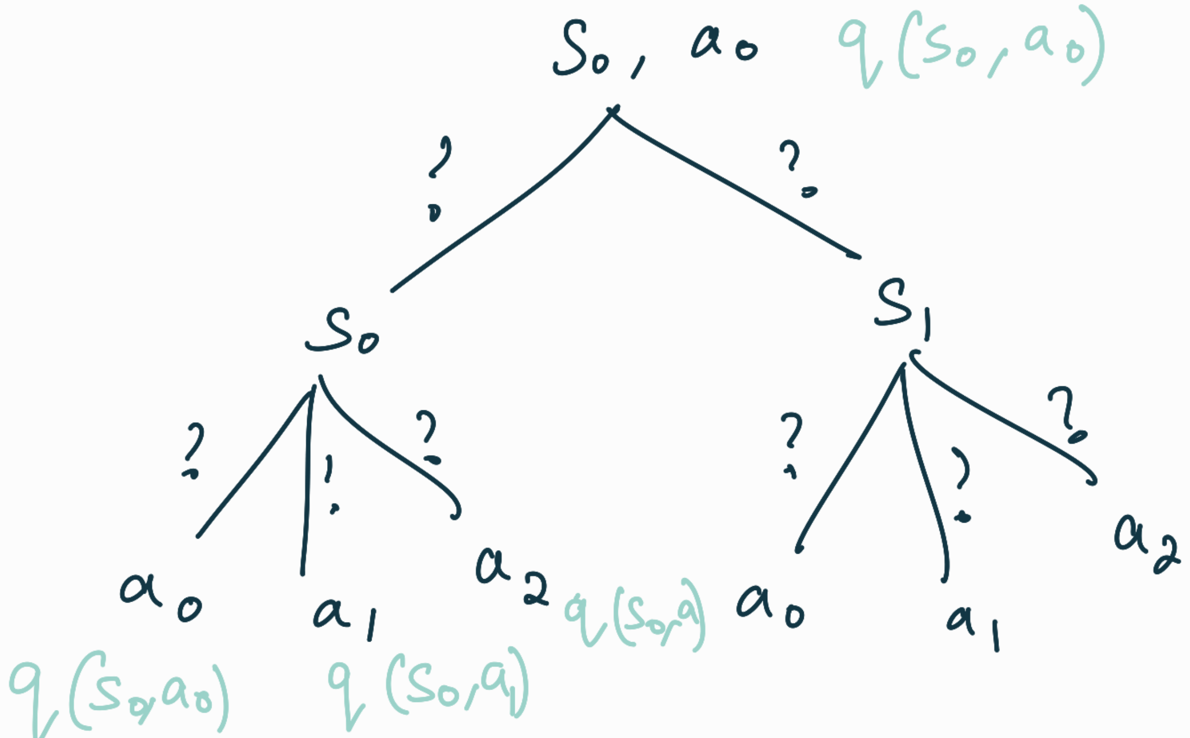
$$v_\pi(s_0) = \sum_{a \in A} \pi(a|s_0)\, q_\pi(s_0, a)$$

$$= \sum_{a \in A} \pi(a|s_0)\left[P(s_0|s_0,a)\left(r(s_0,a,s_0) + \gamma\, v_\pi(s_0)\right) + P(s_1|s_0,a)\left(r(s_0,a,s_1) + \gamma\, v_\pi(s_1)\right)\right]$$

$$v_\pi(s_0) = \sum_{a \in A} \pi(a|s_0) \left(\sum_{s' \in S} p(s'|s_0,a) \left[r(s_0,a,s') + \gamma v_\pi(s') \right] \right)$$

Bellman's equation for state value function.

* s_0, a_0



$$q_\pi(s_0, a_0) = \sum_{s' \in S} p(s'|s_0,a_0) \left[r(s_0,a_0,s') + \gamma \sum_{a' \in A} \pi(a'|s') q_\pi(s',a') \right]$$

$$\left.\begin{aligned} q_*(s_0, a_0) &= 0.7 \\ q_*(s_0, a_1) &= 0.8 \\ q_*(s_0, a_2) &= 0.2 \end{aligned}\right\} \quad \begin{aligned} \pi(a_0|s_0) &= 0 \\ \pi(a_1|s_0) &= 1 \\ \pi(a_2|s_0) &= 0 \end{aligned}$$

$$v_{\pi^*}(s) \longrightarrow \pi^*(a|s)$$

$$v_{OLD}(s) = 0$$

① $v_{NEW}(s) = \sum_{a \in A} \pi(a|s) \left[\sum_{s' \in S} p(s'|s,a) \left(r(s,a,s') + \gamma v_{OLD}(s') \right) \right]$

② $\Delta = \| v_{OLD}(s) - v_{NEW}(s) \|$

③ $v_{OLD}(s) = v_{NEW}(s)$

repeat ① and ②
until $\Delta \leq tol$

$tol = 0.0001$
