Find $\sigma : S \to A$ that maximises :

s_0 initial state

$$Val_\sigma(s_0) = \mathbb{E}_{\sigma, s_0}\left[\sum_{t=0} r_t\right]$$

Value

$Val_\sigma : S \to \mathbb{R}$

$s \in S$

$$Val_\sigma(s) = \mathbb{E}_{\sigma, s}\left[\sum_{t=0}^{\infty} \gamma^t r_t\right]$$

Val_σ is called On-policy value function (σ)

$$\mathrm{Val}_*(s) = \max_{\sigma : S \to A} \mathrm{Val}_\sigma(s)$$

Val_* is called optimal value function

$\mathrm{val}_\sigma, \mathrm{val}_* : S \to \mathbb{R}$

v. values

q. values (Q-learning)

$v : S \to \mathbb{R}$: $v(s)$: total reward from s

$q : S \times A \to \mathbb{R}$: $q(s,a)$: total reward from s playing a

$s \in S$ $a \in A$ $\sigma : S \to A$

$$q_\sigma(s,a) = \mathbb{E}_{\sigma, s, a}\left[\sum_{t=0}^{\infty} \gamma^t r_t \right]$$

$$\mathrm{val}_\sigma(s) = \mathbb{E}_{\sigma, s}\left[\sum_{t=0}^{\infty} \gamma^t r_t \right]$$

$$q_*(s,a) = \max_{\sigma : S \to A} q_\sigma(s,a)$$

<u>Obj</u>:

- Find σ $\max$ $\mathrm{val}_\sigma(s_0)$
- $\mathrm{val}_*(s_0)$

equivalent

- $q_*(s_0, a)$ for all $a \in A$

<u>TAXONOMY</u> OF RL algorithms

Disclaimer: Does not exist

model-free

model-based

know OR learn the MDP

policy gradients approaches

Q-learning approaches

planning

MCTS
Monte Carlo Tree Search

S finite and A finite

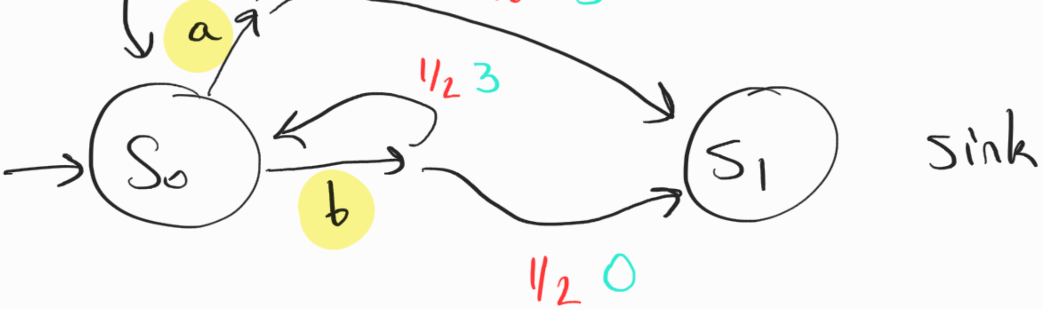
Dynamic setting / Tabular setting

tabular setting

adapt to general case
"function approximation"
using deep RL

Bellman equations :

probability
reward



$$\gamma = 1$$

$$val_*(s_0) = \max \begin{cases} \frac{9}{10}(1 + val_*(s_0)) + \frac{1}{10}(-5+0) & a \\ \frac{1}{2}(3 + val_*(s_0)) + \frac{1}{2}(0+0) & b \end{cases}$$

Bellman equation for a fixed strategy:

$$val_\sigma(s) = \sum_{s', r} \Delta(s, \sigma(s))(s', r)\left[r + \gamma\, val_\sigma(s')\right]$$

Bellman equation for optimal strategy

$$val_*(s) = \max_{a \in A} \sum_{s', r} \Delta(s, \sigma(s))(s', r)\left[r + \gamma\, val_*(s')\right]$$