MDP

S : {states}

A : {Actions}

T : transition function $T(s_t, a, s_{t+1}) = Pr(s_{t+1} | s_t, a)$ PDF over states at time t+1

R : reward function

↳ Simple case: R(S) fixed for a given state

↳ Complex case: R(s, a, s')

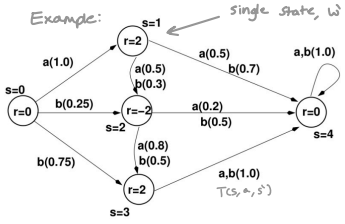
At each time step t, agent is in some state, s_t and must take an action, a_t.
Each action causes a transition to a new state, s_{t+1}.

Representations:

S-A-R Space:

more compact ↓

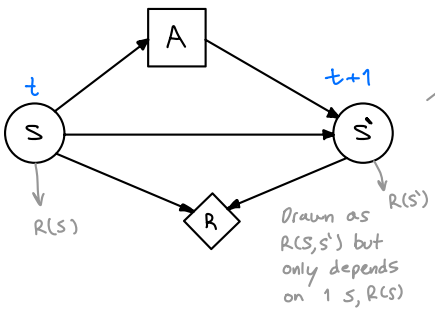
Example:



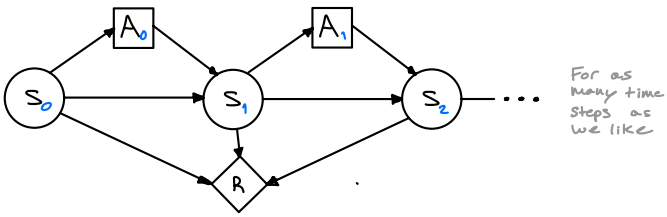
DBN:

* technically a DDN, since R & A aren't R.V's

Compact: represent each State w/ only 2 time slices



Unrolled (Full DBN) →



Optimal policy π^*, gives max **expected reward**. For $t \to \infty$: $\sum_{t=0}^{\infty} \gamma^t R(s_t)$

"Value" of being in S w/ t stages to go, $V^t(S)$.

↳ Find w/ DP: for all s' what action gets us to best-next state?

Start w/ $V^0(S) = R(S)$

Then, w/ t-stages to go → (in practice, until V^t stops changing much; ∞ stage to go) → Optimal

$$V^t(s) = \max_a \left[R(S) + \gamma \sum_{s'} Pr(s'|s,a) V^{t-1}(s') \right] \qquad V^*(s)$$

$$\pi^t(s) = \arg\max_a V^t(s) \qquad \pi^*(s)$$