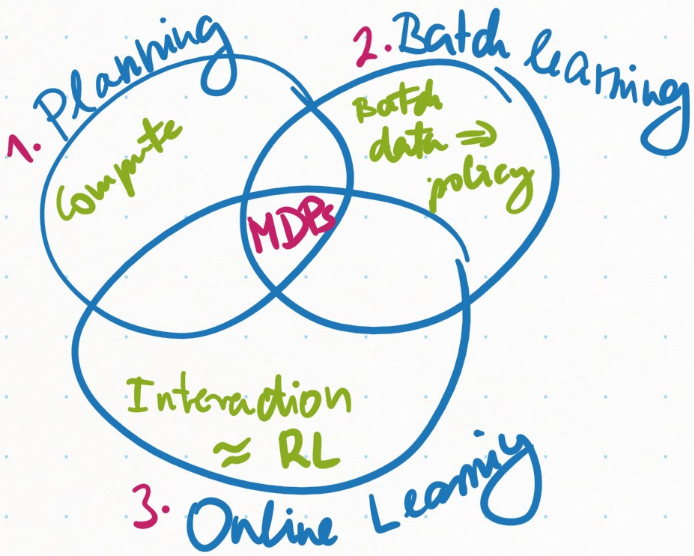
3 Blocks

Block $\approx$ 4 weeks



$\mathbb{R}$ = reals

$0 \leq \gamma < 1$

$\gamma = 1$

return

$\frac{1}{\varepsilon(1-\gamma)}$

MDP = Markov Decision Process

-"- -"- Problem

$M = (S, A, (P_a(s))_{a,s}, (r_a(s))_{a,s})$

States | S

Actions | A

Stochastic transitions (between states)

$s, a \mapsto$ distr. over states

| $P_a(s)$

Rewards | $r_a(s) \in \mathbb{R}$ $(\in [0,1])$

Objective

$\tau = (S_0, A_0, S_1, A_1, \ldots\ldots)$

$R(\tau) = r_{A_0}(S_0) + \gamma r_{A_1}(S_1) + \gamma^2 r_{A_2}(S_2) + \ldots$ → max.

$\mathcal{M}_1(X) \equiv$ set of prob. distr. over X

Policies!

$$H_t = \underbrace{(S_0, A_0, \dots, S_{t-1}, A_{t-1}, S_t)}_{\downarrow} \in \overbrace{(S \times A)^{t} \times S}^{\mathcal{H}_t}$$

$$\mathcal{M}_1(A)$$

$$\boxed{\begin{array}{l} \pi = (\pi_t)_{t \geq 0} \\ \pi_t : \mathcal{H}_t \to \mathcal{M}_1(\mathcal{A}) \end{array}}$$

State is observed!?!

$S, \mathcal{A}$: measures on them??

S : finite

$\mathcal{A}$: finite

BIG!

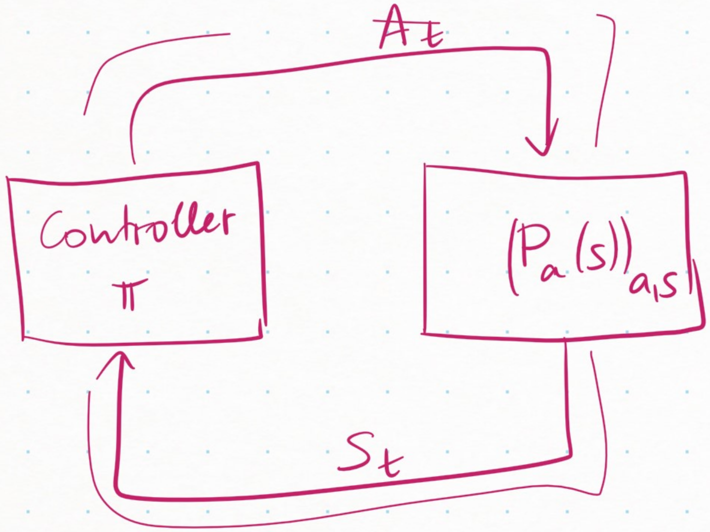
$$\pi_t(a \mid h_t)$$

$$\downarrow$$

$$h_t \mapsto \pi_t(U \mid h_t)$$

$$\boxed{U} \subseteq \mathcal{A}$$

High bar!



"Closed-loop" / Feedback

$$\overbrace{\pi, \mu \in M_1(S), \quad (P_a(s))_{s,a}}^{\text{interconnection}}$$

$$T = (S\times A)^{\mathbb{N}} = \{(s_t,a_t)_{t\geq 0} \mid s_t\in S,\ a_t\in A\}$$

$$\mathbb{P}^\pi_\mu((s_0,a_0,s_1,a_1,\dots)) = \dots$$

$$= \mu(s_0)\times\underbrace{\underbrace{\pi_0}(s_0)(a_0)}_{\substack{\pi_0(a_0|s_0)\\ \pi(a_0|s_0)}}\times\underbrace{(P_{a_0}(s_0))(s_1)}_{\substack{P_{a_0}(s_0,s_1)\\ P_{a_0}(s_1|s_0)}}$$

$$\times\underbrace{\pi_1(s_0,a_0,s_1)(a_1)}_{\substack{\pi_1(a_1|s_0a_0s_1)\\ \pi(a_1|s_0a_0s_1)}}\times\underbrace{(P_{a_1}(s_1))(s_2)}_{P_{a_1}(s_1,s_2)}$$

$\vdots$

"Markov property"

$$v^\pi(\mu) = \mathbb{E}^\pi_\mu[R(\tau)] \doteq \sum_{g\in T}\mathbb{P}^\pi_\mu(g)R(g) \quad \Big|\ v^\pi_M(\mu)$$

value of π in MDP M

$$\sup_{\pi} v^{\pi}(\mu) = v^{*}(\mu)$$

$$\mu = \delta_s \qquad \delta_s(s') = \begin{cases} 1, & s = s' \\ 0, & \text{o.w.} \end{cases}$$

$$v^{\pi}(\delta_s) =: v^{\pi}(s)$$

Homework: Figure out relationship between $v^{\pi}(\mu)$ and $v^{\pi}(s)$

$$v^{*}(s) = \sup_{\pi} v^{\pi}(s)$$

Examples?

Fundamental Theorem of MDPs