Reinforcement Learning

Final exam : Friday May 15th

- 7:15 PM - 10:15 PM
- Online
- Multiple choice
- Calculator needed
- Practice Gradiance Quiz will be posted
- One more Gradiance Quiz left (RL)

$\mathcal{A}$ - set of all actions

$A_t \in \mathcal{A}(S_t) \subset \mathcal{A}$

$R_t \in \mathbb{R}$

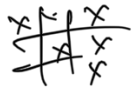
Policy π

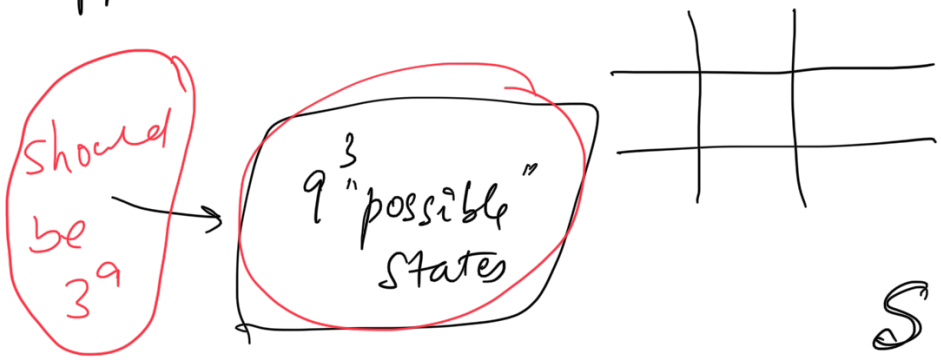
$S_1 \rightarrow A_{S_1}$

$S_2 \rightarrow S_2$

TicTacToe

Playing against an inperfect opponent.

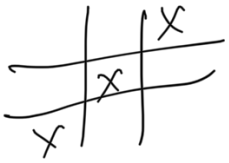


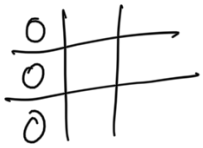


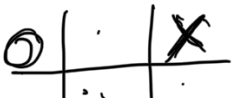
Value Prob of winning from that state.

S is small for ia.


0.5


1


0



$A(S_0)$

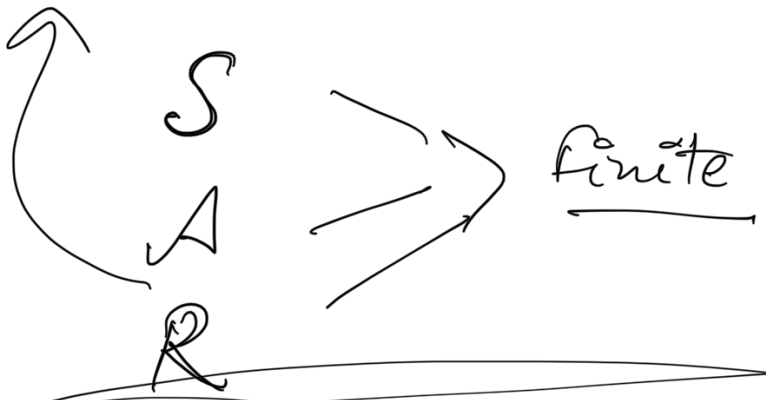
S_1

$0.5 + \alpha(0 - 0.5)$

$0.5 - 0.5\alpha$

Markov Decision processes (MDP)

Finite MDP



$$p(s', r \mid s, a) \doteq P(S_t = s', R_t = r \mid S_{t-1} = s, A_{t-1} = a)$$

$|S| = 2 =$ on, off

$|A| = 4 = u, d, l, r$

$|R| = 2 = +1, -1$

$p(on, +1 \mid on, u)$

→ $p(s', r \mid s, a)$

$$p(s' \mid s, a) \doteq \sum_{r \in R} p(s', r \mid s, a)$$

G_t = Expected return from $t \rightarrow T$