Lec-4

14th Sept 2017

- Concepts
  - MDP : $M = (S, A, P, R, d_0, \gamma)$
  - Discrete / Continuous
  - Policy / Optimal policy
  - History / episode / trajectory
  - Terminal state
  - Planning vs RL vs
  - Markov property
    - Markovian state representation
  - Stochastic vs deterministic
  - stationary vs Non-stationary
  - Partial observability

Stationary vs Non-stationary

- P stationary unless otherwise defined

$$Pr(S_{t+1} = s' \mid S_t = s, A_t = A) = Pr(S_{i+1} = s' \mid S_i = s, A_i = A)$$

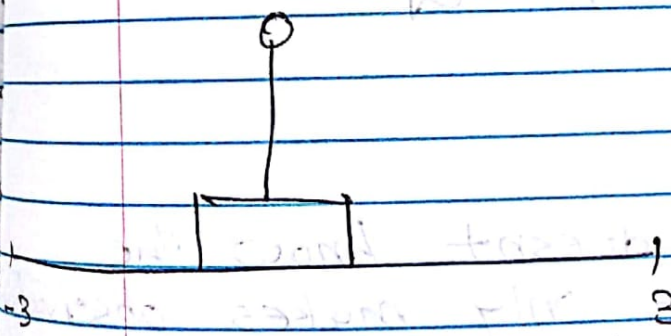
- Reward f^n is stationary unless otherwise defined

$$Pr(R_t = r \mid S_t = s, A_t = a, S_{t+1} = s') = Pr(R_i = r \mid S_i = s, A_i = a, S_{i+1} = s')$$

- Similarly for policy

NOTE:

- Cart pole



State : $(\theta, x, v, \dot{\theta})$

S_0 = Vertical / centered

Action : (left, Right)

$R_t = 1$, always

if pole falls = end of episode

OR episode terminates in 20 sec.

hence, we have to augment our state set with time,

$\therefore S = (\theta, x, v, \dot{\theta}, t)$

- Finite horizon MDP

$\exists L, \forall t \geq L, s_t = s^*$

- Horizons —

  - Finite —
  - Indefinite - all episodes will terminate in finite time but cannot specify L.
  - Infinite — $L = \infty$

- Partial Observability

  - When agent doesn't know the true state, it only makes observations about the state
  - We can have MDPs with markovian states & non-Markovian observations