

687 2017-09-14

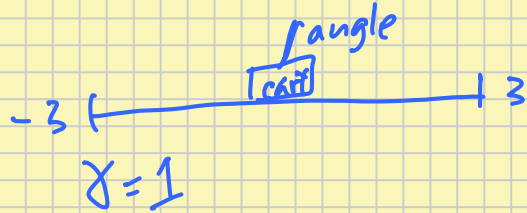
Stationary vs Non-stationary

$$Pr(S_{t+1} = s' | S_t = s \wedge A_t = A) = Pr(S_{i+1} = s' | S_i = s \wedge A_i = A) \quad \forall i$$

Similarly for rewards R . Policy may vary (learning!)

What about variation over ^{multiple} episodes? Environment may shift in some problems.

Cart-Pole Problem (Inverted Pendulum)



$$\text{state} = (\theta, x, \dot{\theta}, \dot{x})$$

$$s_0 = (0, 0, 0, 0) \quad \text{actions (L, R)}$$

episode ends if hit end of track.

or if pole falls ($> 45^\circ$)

or at $t \geq 20 \text{ sets}$ $\Delta t = .02s$

accelerate (force) ^{apply}

$R_t = 1$ always
(max time balanced)

Not Markovian! Why?

→ This constraint requires we know the time.

$$\hookrightarrow \text{State} = (\theta, x, \dot{\theta}, \dot{x}, t)$$

A case of a Finite Horizon MDP:

$$\exists L: \forall t \geq L: S_t = \bar{s}$$

↖ the horizon. Add t to state space (perhaps implicitly).

Can also have Indefinite Horizon - all episodes terminate, but no bound \angle
Infinite Horizon - some episodes may never terminate.

Partial Observability:

- Agent does not know true state - only has observations
- Can be noisy, incomplete.
- Can have Markovian states + non-Markovian observations
- Two approaches:

POMDP: state \rightarrow sensor \rightarrow agent

Function Approximation: state \rightarrow agent but observations modeled in agent

\uparrow Used in this course.

Any system can be modeled either way.

Affects whether environment is an MDP (Markovian).

\downarrow
Func. Approx gives simpler model of the environment.