Reinforcement Learning

Saurabh Gupta

Agent Environment Interface

Reinforcement Learning

Image Source: https://www.freecodecamp.org/news/a-brief-introduction-to-reinforcement-learning-7799af5840db/

Markov Decision Process

o_t

Step Back

a_t

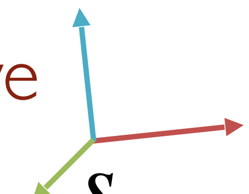
o_{t+1}

$\dots$

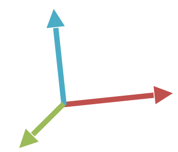
Transition Function

How you move,
how the tiger moves?

3D Relative
Pose



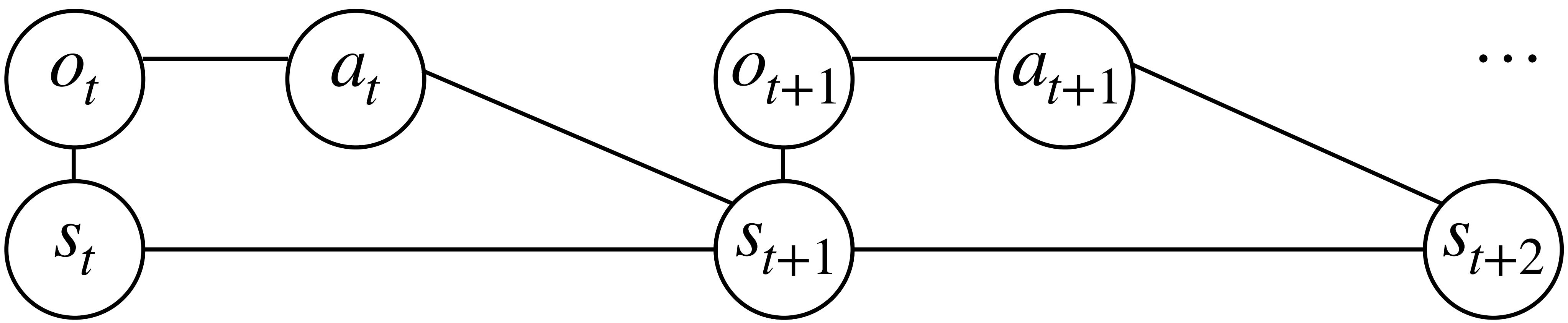
s_t



s_{t+1}

Reward Function

Survived?



One step dynamics $p(s_{t+1}, r_{t+1} \mid s_t, a_t)$

Transition Function $p(s_{t+1} \mid s_t, a_t)$ $p(s_{t+2} \mid s_{t+1}, a_{t+1})$

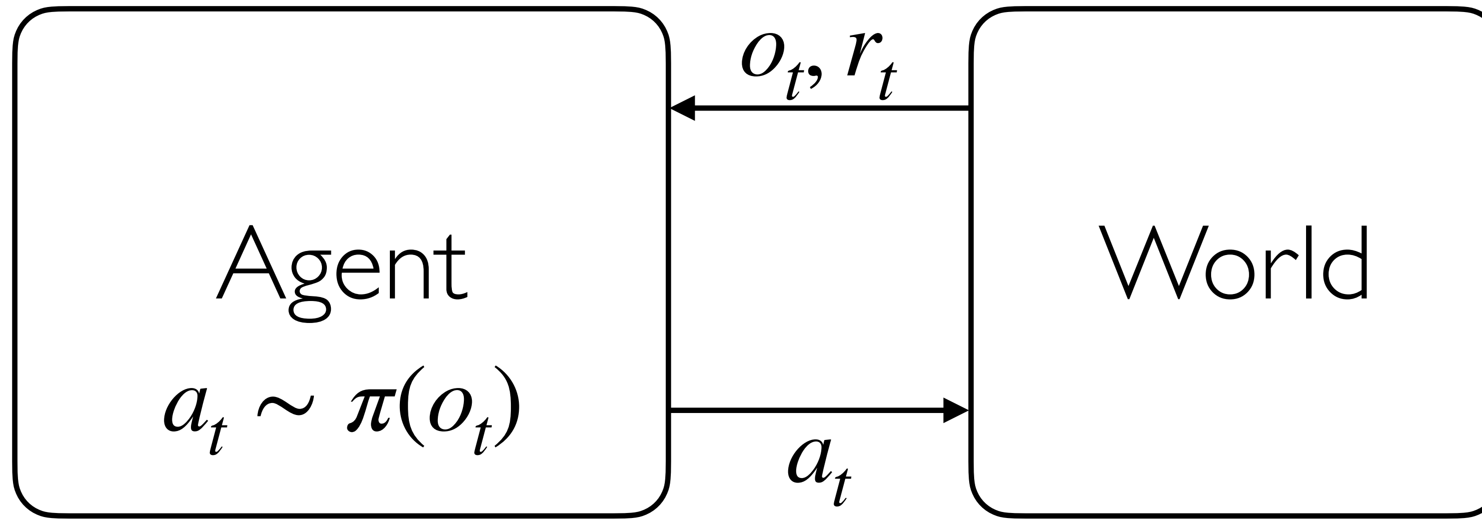
Reward Function $r_{t+1} = R(s_{t+1}, s_t, a_t)$ $r_{t+2} = R(s_{t+2}, s_{t+1}, a_{t+1})$

Goal $\text{argmax}_{a_0, \dots, a_T} \sum_t \gamma^t r_t$

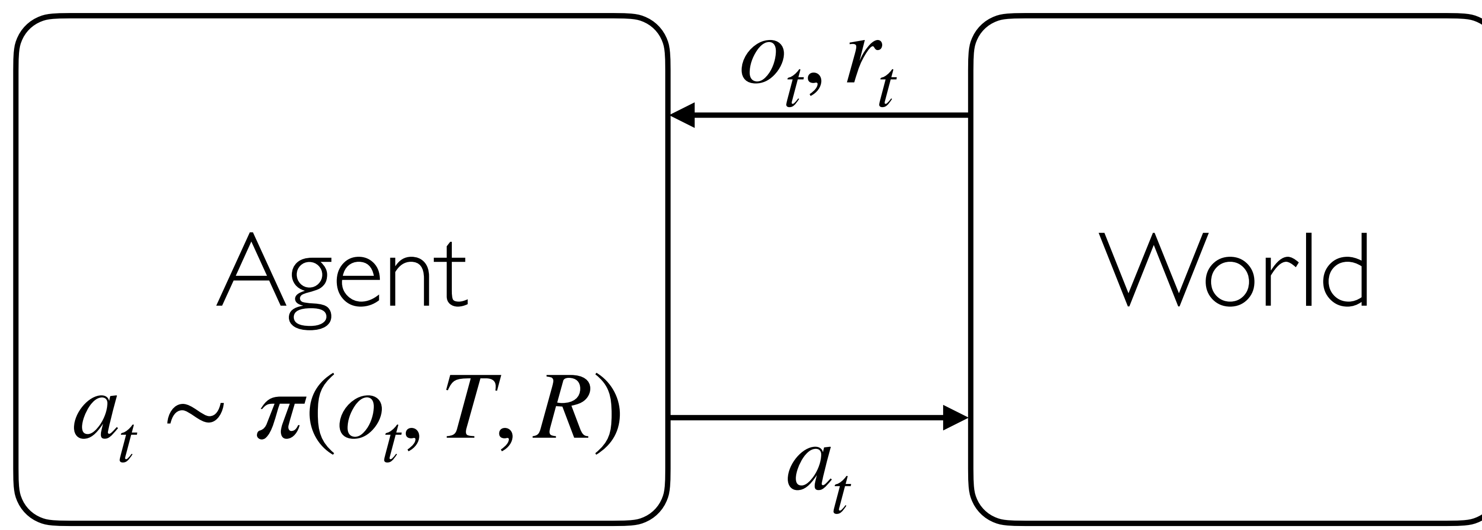
Solving MDPs

Policy: $a_t \sim \pi(o_t)$

Most General Case



More Specific Case



Fully Observed System $o_t = s_t$

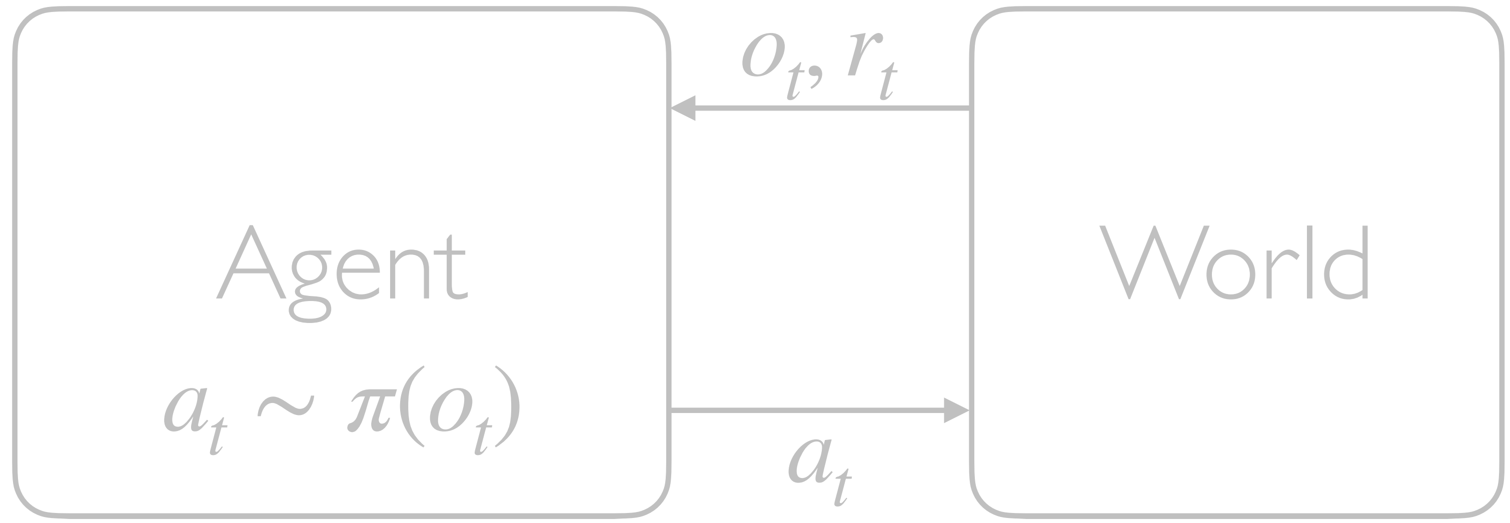
Known Transition Function $s_{t+1} \sim T(s_t, a_t)$

Known Reward Function $R(s_{t+1}, s_t, a_t)$

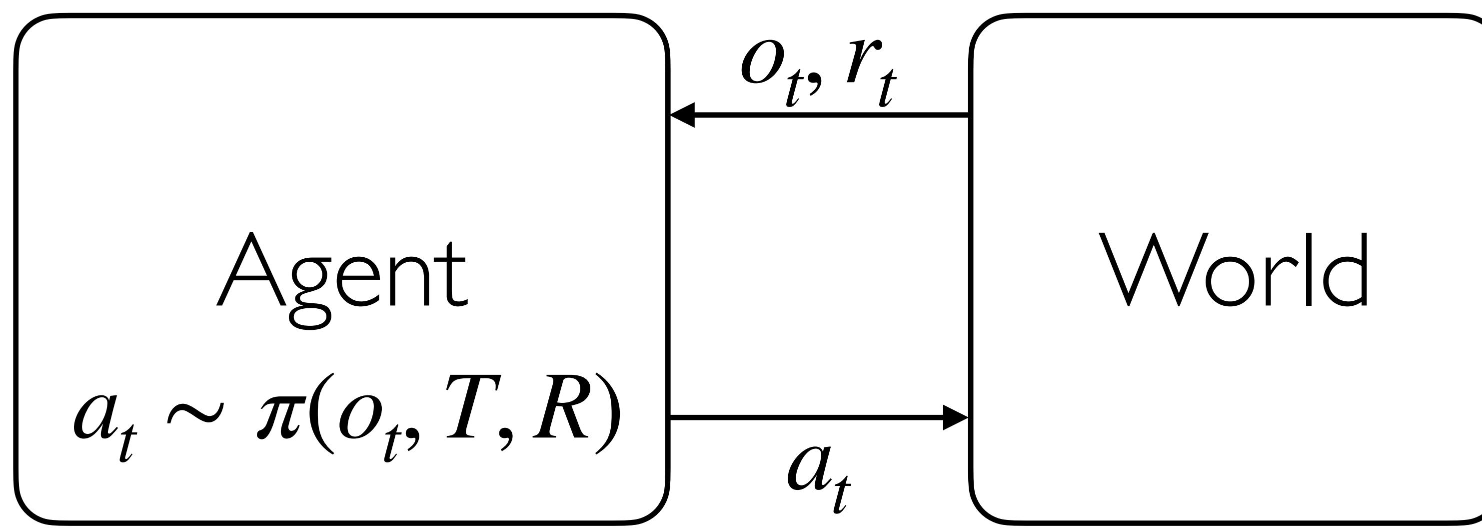
Solving MDPs

Policy: $a_t \sim \pi(o_t)$

Most General Case



More Specific Case



Fully Observed System	$o_t = s_t$
Known Transition Function	$s_{t+1} \sim T(s_t, a_t)$
Known Reward Function	$R(s_{t+1}, s_t, a_t)$

Basics

Policy

Episodes

Returns

Value Functions

Action-value Functions

Solving MDPs via Dynamic Programming

Policy Evaluation

Policy Improvement

Policy Iteration

Value Iteration

Resources

Reinforcement Learning: An Introduction
Sutton and Barto
http://incompleteideas.net/book/the-book-2nd.html

David Silver's **Reinforcement Learning Course**
https://www.davidsilver.uk/teaching/

Thank you