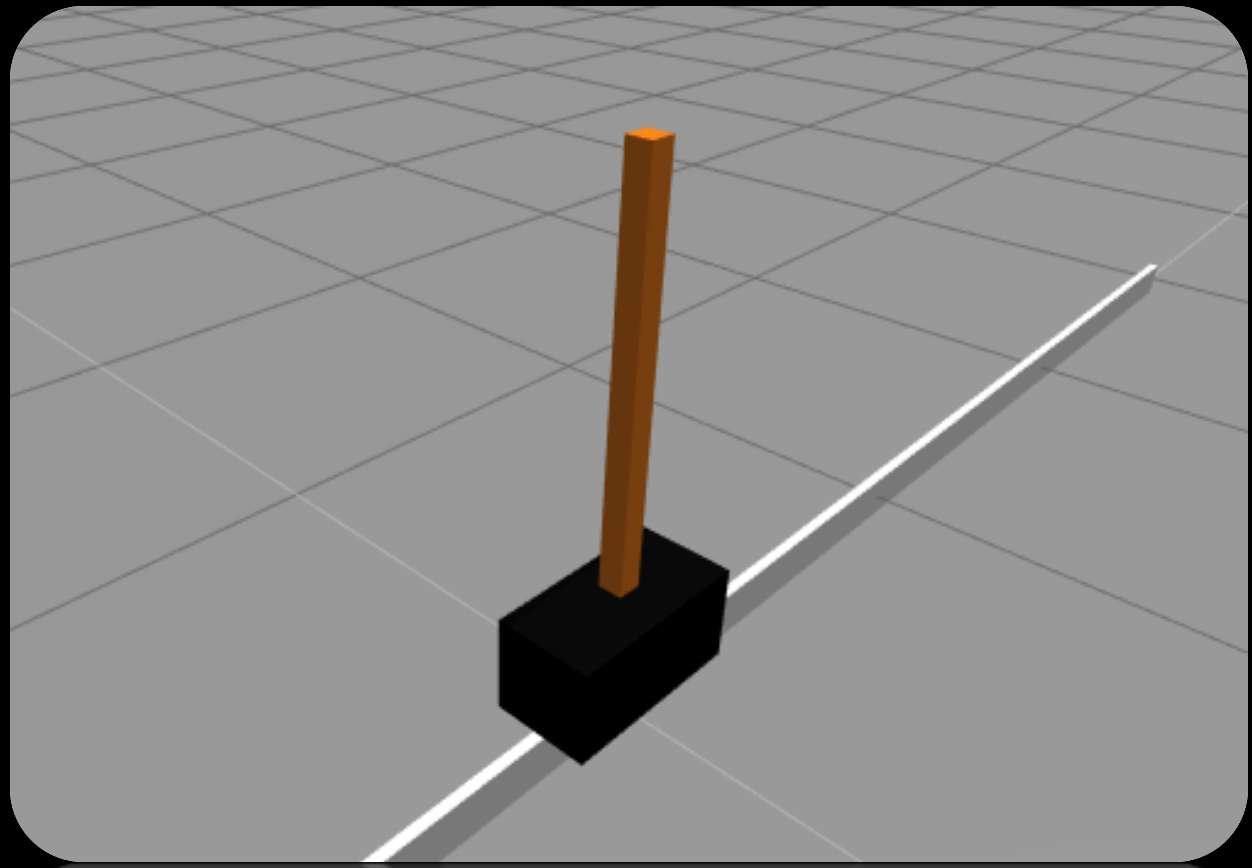
THE BUILDING CULTURE

# ROBOTICS & AI MEETUP #12

JOIN US ON

DISCORD

7 PM CET, 20TH MARCH



## CART POLE BALANCE USING Q-LEARNING

# MEETUP #12 : 20/03/2021

THE BUILDING CULTURE

## AGENDA

19:00 Introductory Presentation

19:05 Project Presentation

19:45 Discussion

20:00 Feedback

20:05 Introductions

20:30 Casual Networking

# WHAT IS THE BUILDING CULTURE COMMUNITY ?


## ROBOTICS EDUCATION

Best way to learn is to build

ACTIVE LEARNING GROUP OF ROBOTICS & AI ENTHUSIASTS

# PLEASE FILL THE ROLES

THE BUILDING CULTURE

- Project presenter (*#meetup-volunteer* in Events)
- Research Paper presenter
  (*#paper-reading-volunteer* in Events)
- News-feed updater in all 4 categories

**VOLUNTEER-DRIVEN COMMUNITY !**

MEETINGS WILL BE RECORDED AND SHARED ON YOUTUBE

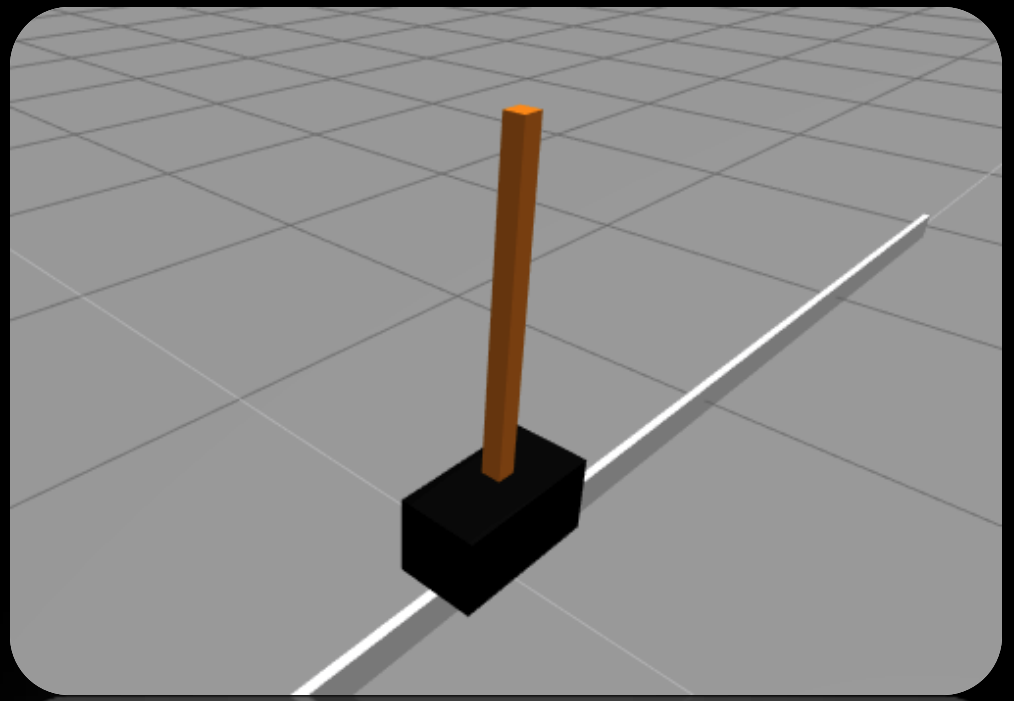
EVERYONE CAN PARTICIPATE REGARDLESS OF EXPERIENCE OR AGE GROUP


# LET'S START

# Robotics & AI Meetup #12

## Cart Pole Balance using Q-Learning

- Goal
- Problem Definition
- Brainstorming
- Project / Case Study Solution
- Algorithm + Code Walkthrough
- Concepts Covered
- Task for next week



BEGINNER FRIENDLY

THE BUILDING CULTURE

# Meetup #11 – RL Concepts


RL PROBLEM

STATE SPACE DISCRETIZATION

EXPECTED RETURN

TEMPORAL DIFF

Q - LEARNING

BELLMAN EQUATION

STATE VALUE FUNCTION

ACTION VALUE FUNCTION

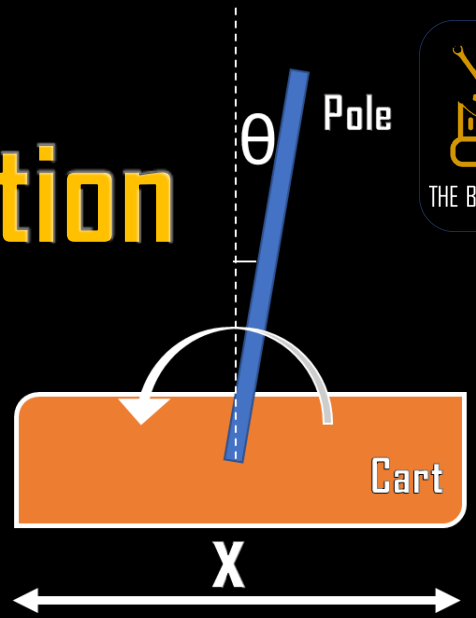
MONTE CARLO METHODS

MODEL BASED RL

MODEL FREE RL

MARKOV DECISION PROCESS

# Meetup #10 – Temporal Diff. Prediction



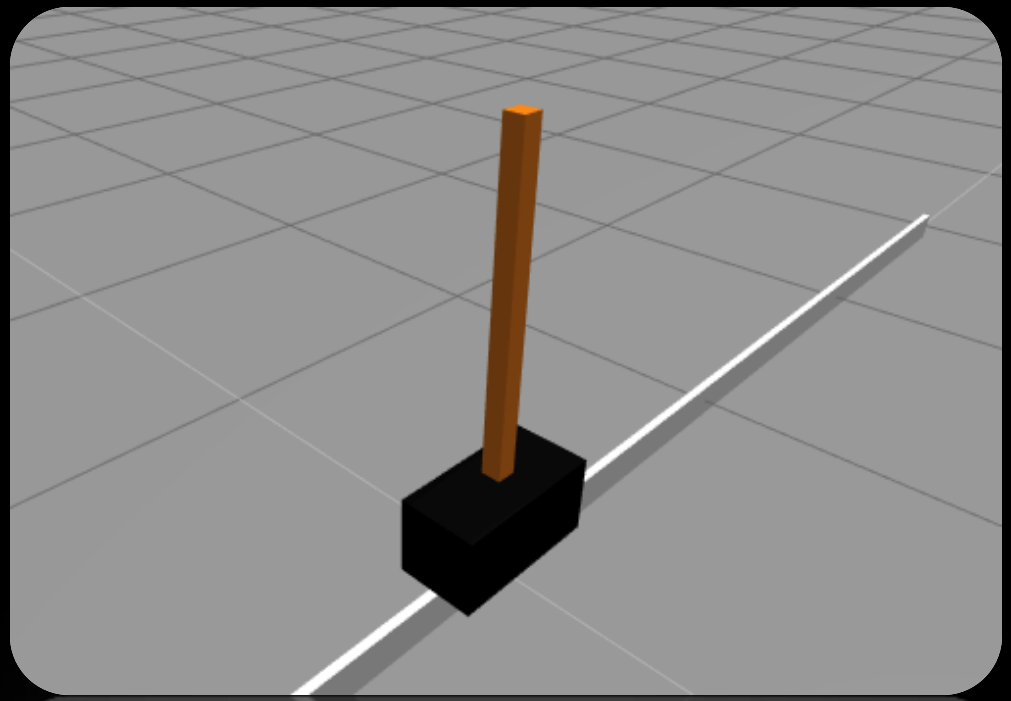

- Initialize the policy
- Initialize value function
- For N no. of episodes :
- initialize environment
- For each step in episode:
- Select action acc to given policy
- Take action
- Get reward, new state
- Set current state to new state

- Discretize pole angles
- If theta < 0, move left
- else move right
- 5000 games, print V for all 10 states to terminal
- Alpha = 0.1, Gamma = 0.99

# Robotics & AI Meetup #12

## Cart Pole Balance using Q-Learning

- Goal
- Problem Definition
- Brainstorming
- Project / Case Study Solution
- Algorithm + Code Walkthrough
- Concepts Covered
- Task for next week



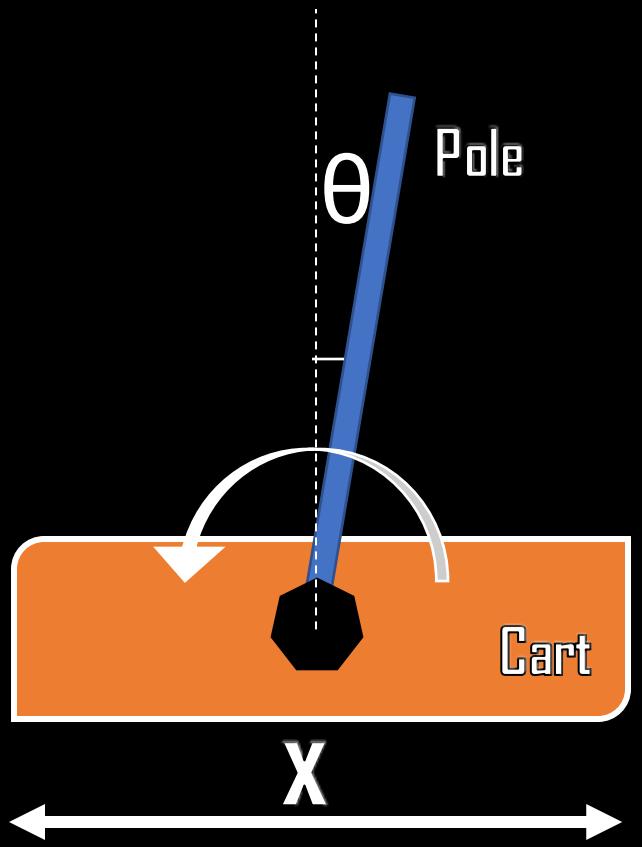
BEGINNER FRIENDLY

THE BUILDING CULTURE

# Problem Definition

## Goal

- Learn the following :
- Demonstrate and understand Q Learning in code
- Revisit all concepts visited so far
- Solidify our foundation

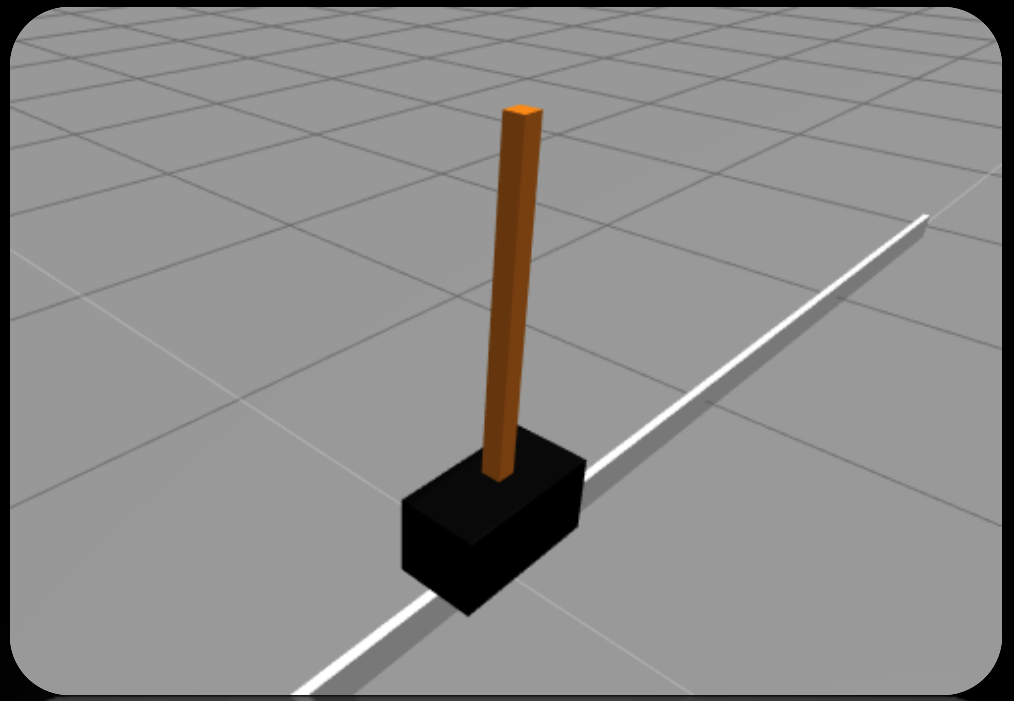



THE BUILDING CULTURE

# Robotics & AI Meetup #12

## Cart Pole Balance using Q-Learning

- Goal
- Problem Definition
- Brainstorming
- Project / Case Study Solution
- Algorithm + Code Walkthrough
- Concepts Covered
- Task for next week

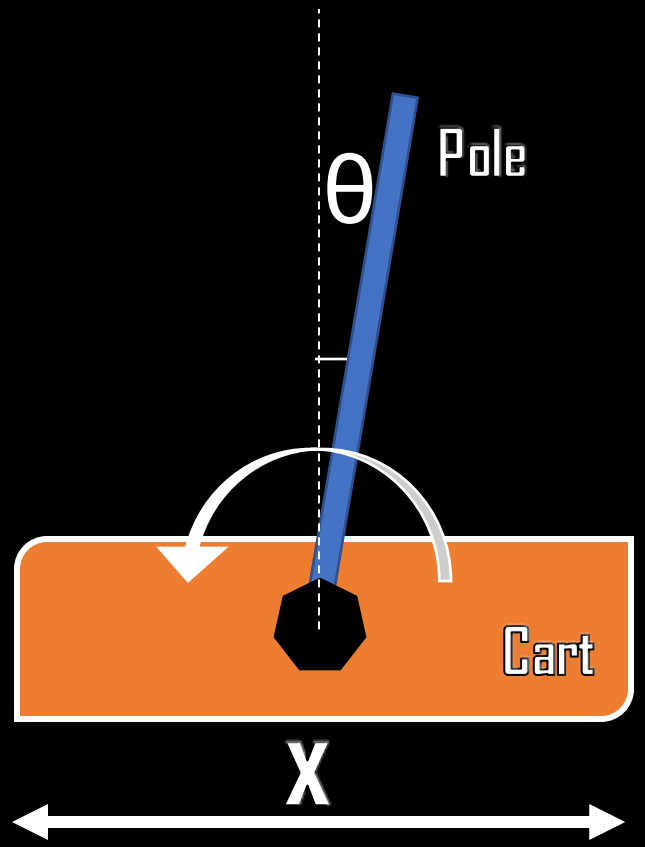


BEGINNER FRIENDLY

# Problem Definition

## Problem

- **Given :**
- Cart moves linearly – x
- Pole rotates - θ
- **Task :**
- Balance pole by moving the cart (*Goal*)
- Balance pole - Keep |θ| <= 12 deg
- Move cart – x (*User controlled*)

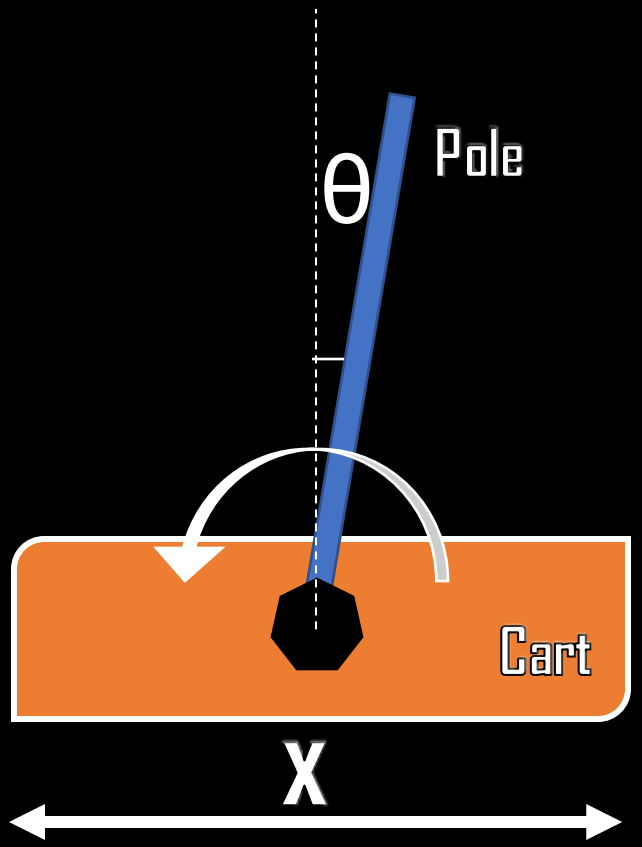


# Problem Definition

## Requirements

- Episode ends when :
- |θ| >= 12 deg – Pole out of balance
- |x| >= 2.4 – Out of range
- Steps >=200

**Balance pole for atleast 200 steps**

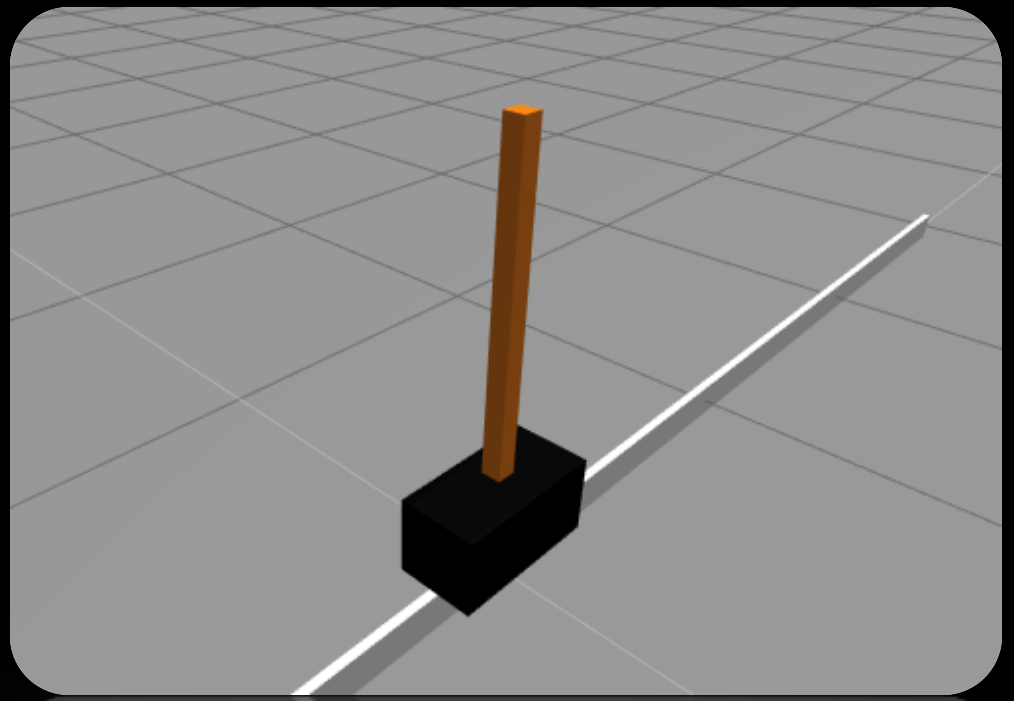



THE BUILDING CULTURE

# Robotics & AI Meetup #12

## Cart Pole Balance using Q-Learning

- Goal
- Problem Definition
- Brainstorming
- Project / Case Study Solution
- Algorithm + Code Walkthrough
- Concepts Covered
- Task for next week



BEGINNER FRIENDLY

THE BUILDING CULTURE

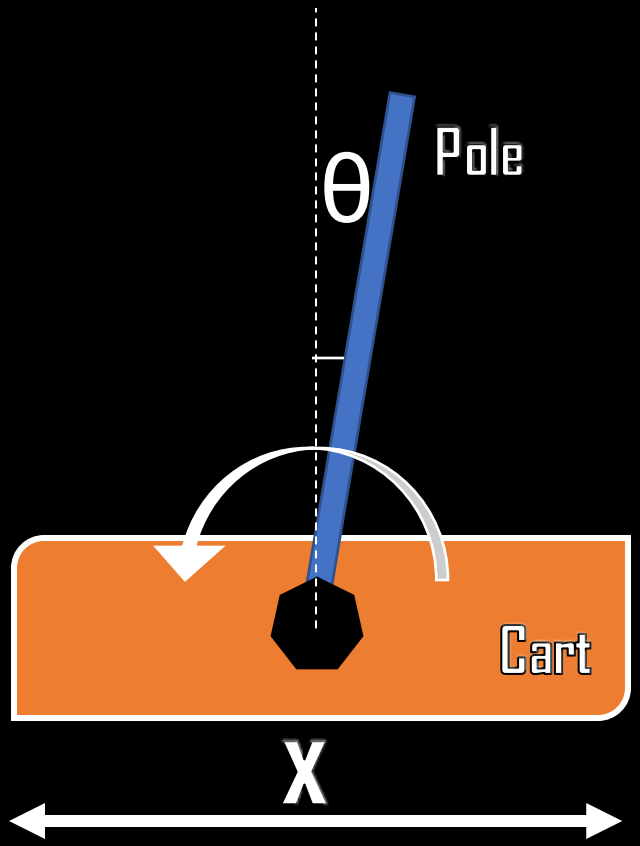
# Problem Definition

State

Action

Reward

Policy




IDEAS ?

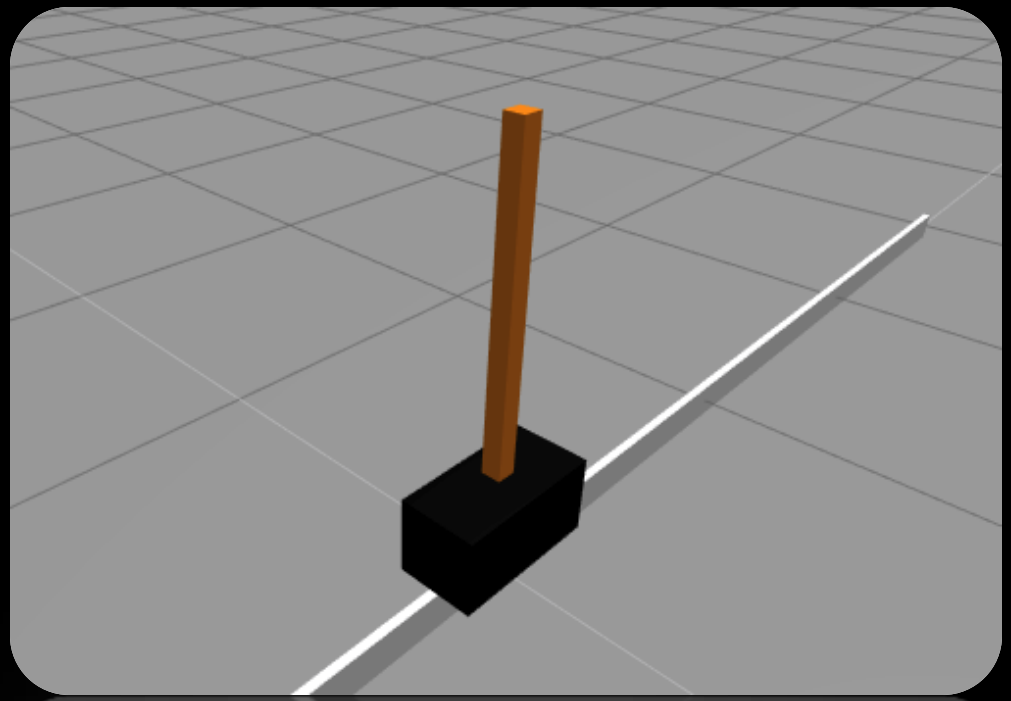
THE BUILDING CULTURE

# Robotics & AI Meetup #10

## Cart Pole Balance using Temporal Difference Prediction

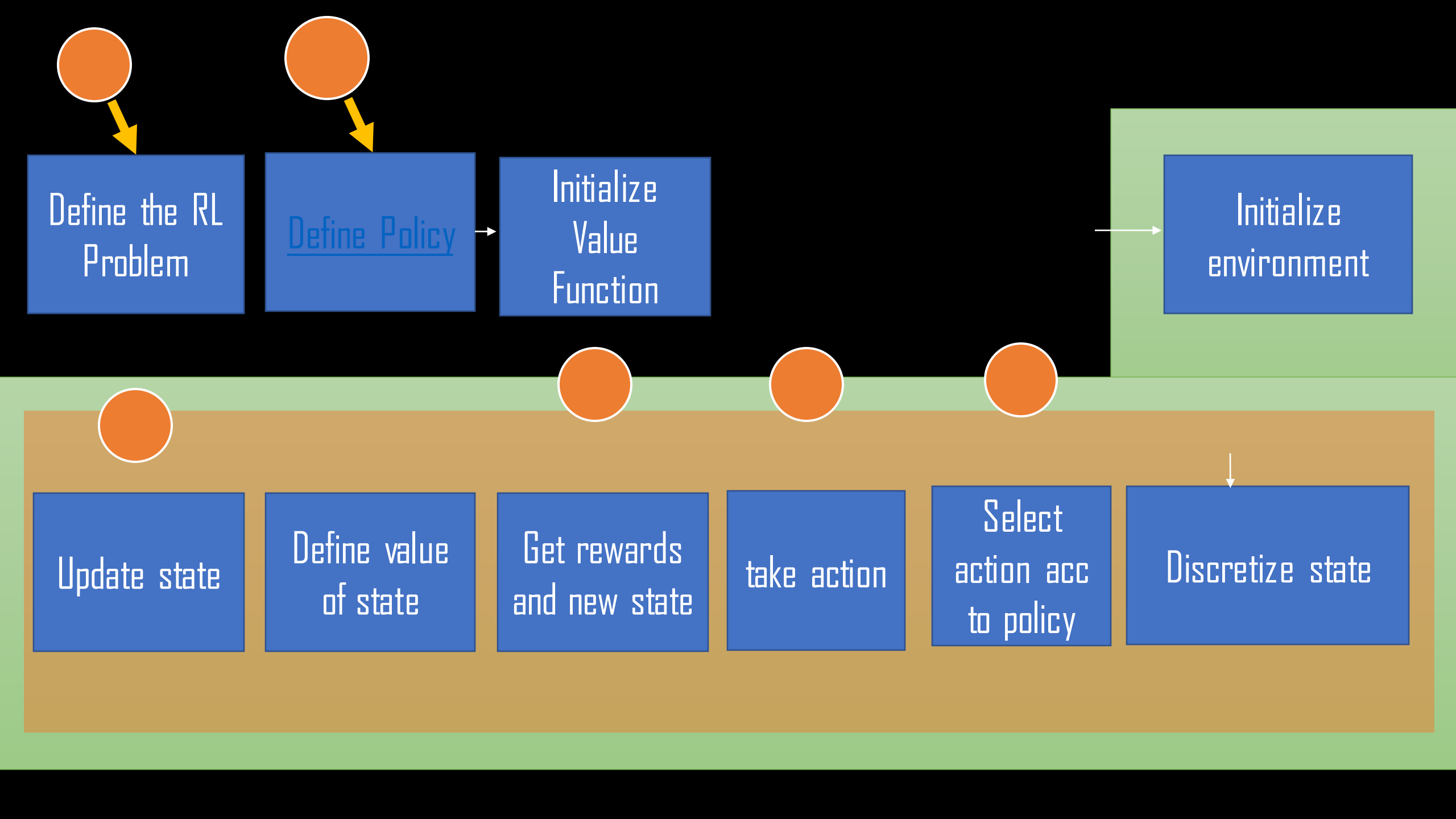
First session

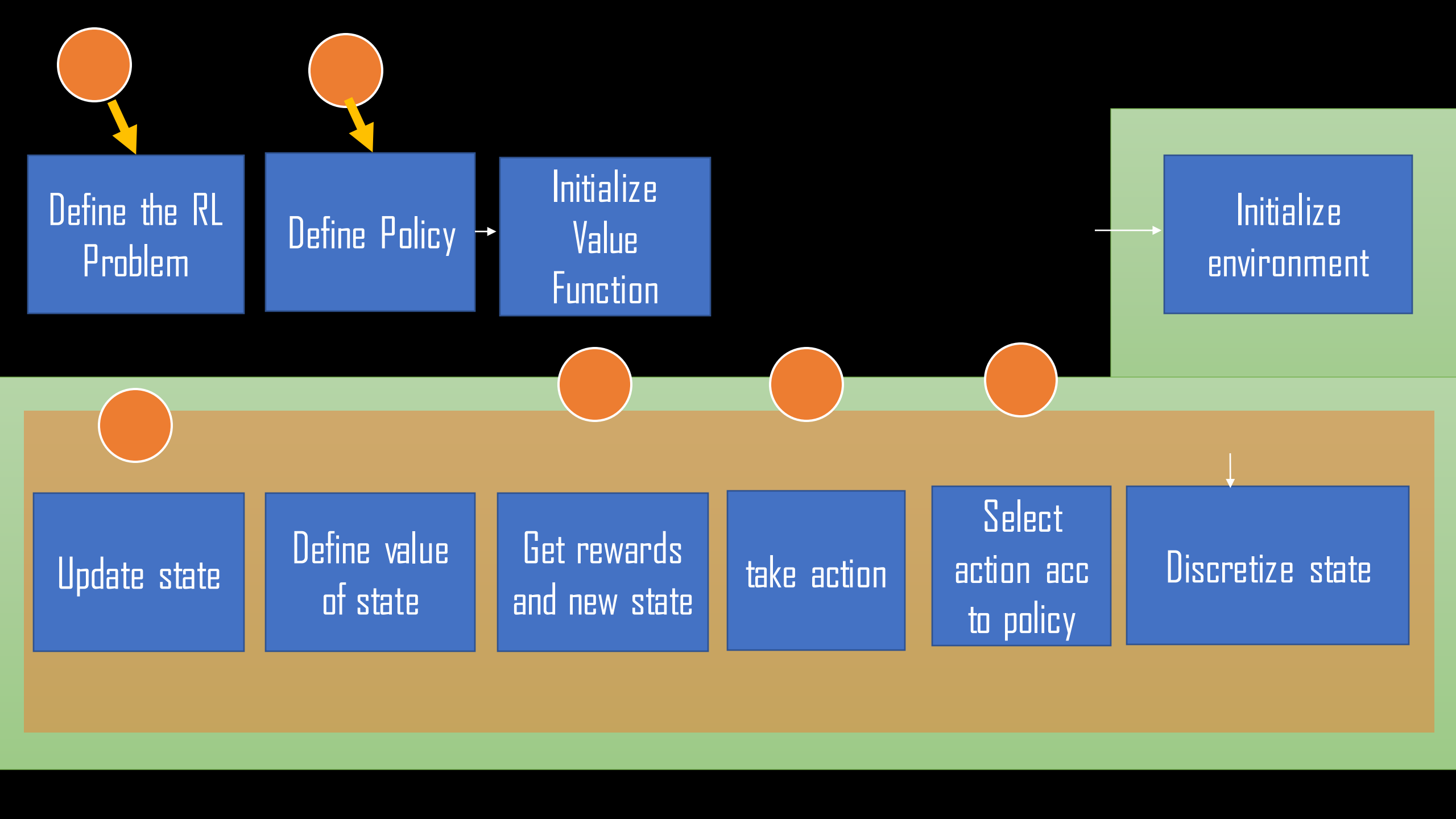
- Goal
- Problem Definition
- Brainstorming
- Project / Case Study Solution
- Algorithm + Code Walkthrough
- Concepts Covered
- Task for next week



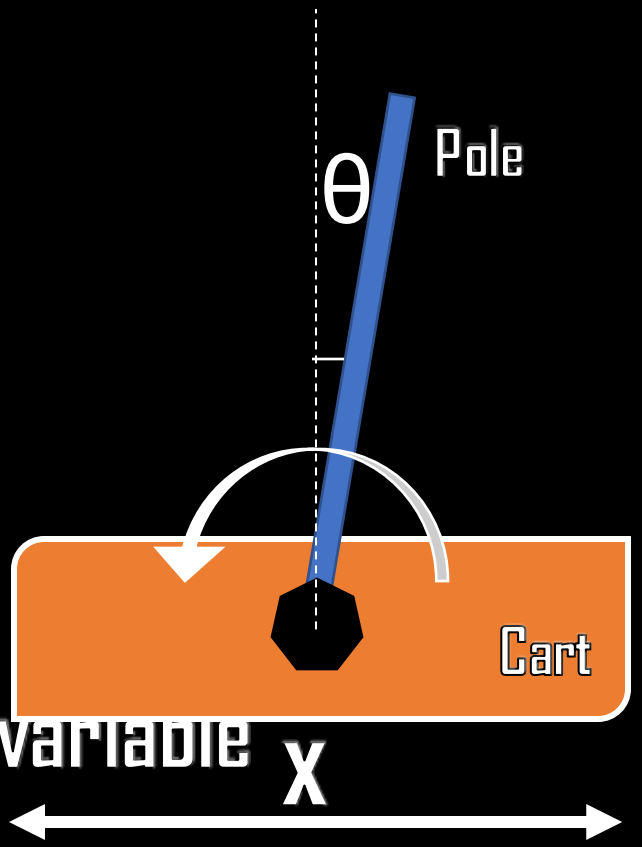
BEGINNER FRIENDLY

THE BUILDING CULTURE


Define the RL Problem
Define Policy
Initialize Value Function
Initialize environment
Update state
Define value of state
Get rewards and new state
take action
Select action acc to policy
Discretize state


Define the RL Problem
Define Policy
Initialize Value Function
Initialize environment
Update state
Define value of state
Get rewards and new state
take action
Select action acc to policy
Discretize state

# Problem Definition




## State

[θ]

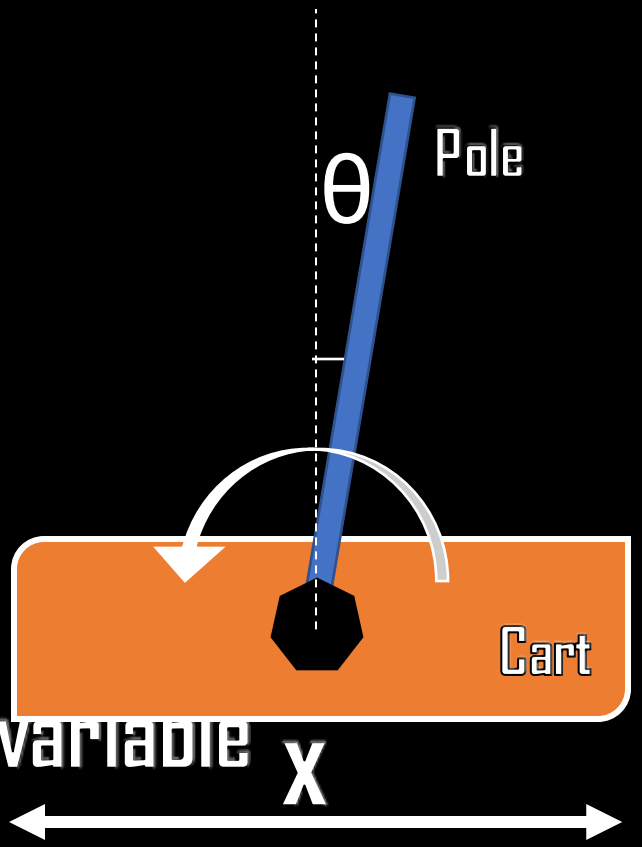
- θ – Pole angle

## State Space

- Range of allowable values for each state variable
- [-0.2019 to 0.2019] radians
- Or [-12 to 12] degrees

CONCEPT : STATE SPACE DISCRETIZATION ⚠

# Problem Definition




## State

[θ]

- θ – Pole angle

## State Space

- Range of allowable values for each state variable
- [-0.2019 to 0.2019] radians
- Or [-12 to 12] degrees

DISCRETE SET OF ANGLES

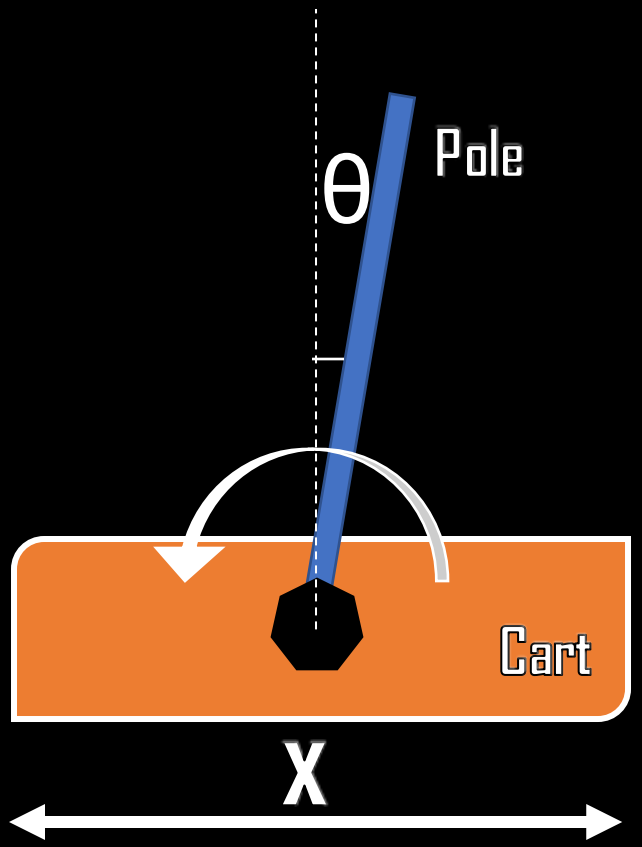
1 2 3 4 5 6 7 8 9 10

-0.2019 +0.2019

New state space

[1,2,3,4,5,6,7,8,9,10]

# Problem Definition




## Action

[0, 1]

- 0 – Moving x left
- 1 – Moving x right

## Reward

R(state, action, next_state) = +1

- Reward of +1 every step when it keeps |θ| <= 12 deg (Pole upright)

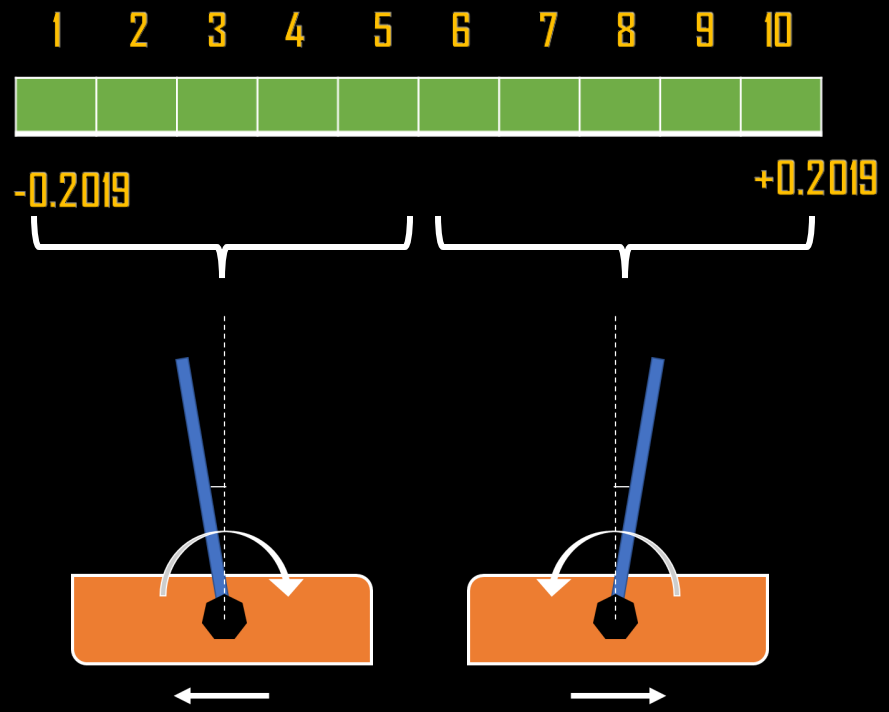
## Goal reached

When running average value of rewards > 195

# Problem Definition



## Policy




- If pole is left of the center
  Move cart left
- If pole is right of the center
  Move cart right

CONCEPT : VALUE FUNCTION ⚠

# Algorithm – Temporal Difference Prediction

- Define the policy
- Initialize value function

[-0.2019 to 0.2019]
10 discrete states

Value[all 10 states] = 0

- If pole is left of the center
  Move cart left
- If pole is right of the center
  Move cart right

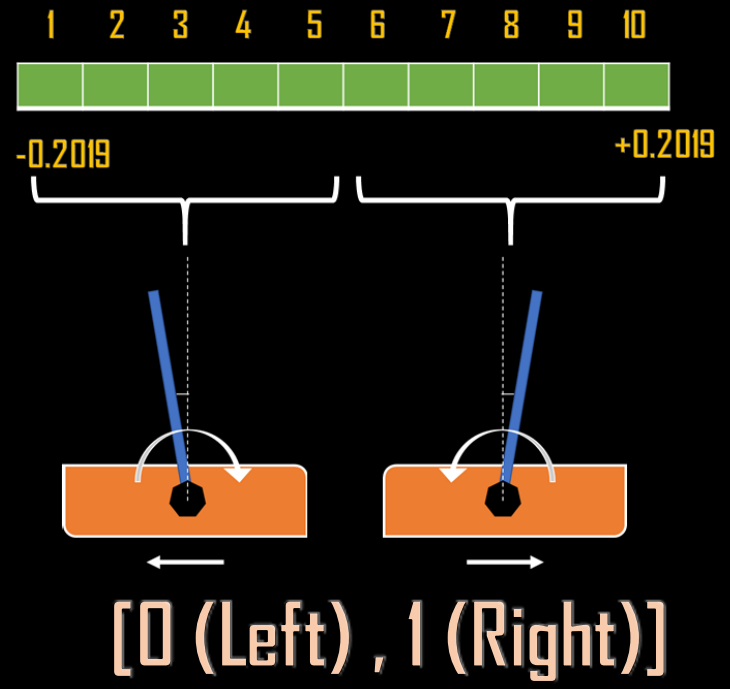
CONCEPT : VALUE FUNCTION ⚠

# Algorithm – Temporal Difference Prediction

- Initialize the policy
- Initialize value function
- For N no. of episodes :
- initialize environment

# Algorithm –Temporal Difference Prediction

- Initialize the policy
- Initialize value function
- For N no. of episodes :
- initialize environment
- For each step in episode:
- Discretize state
- Select action acc to given policy
- Take action



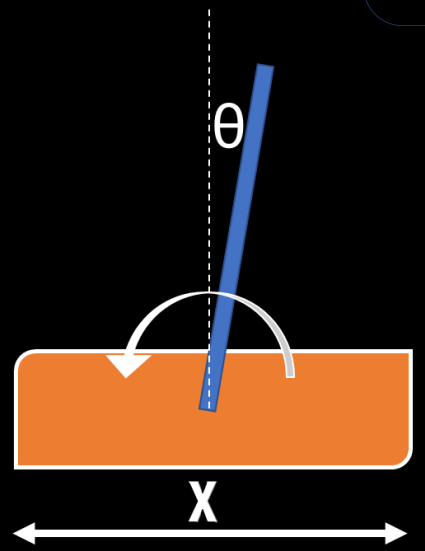
# CONCEPT : MONTE CARLO PREDICTION

## Algorithm (contd.)

THE BUILDING CULTURE

For each step in episode:

- Discretize state
- Select action acc to given policy
- Take action
- Get reward, new state




if |θ| <= 12 deg

+1

# CONCEPT : TEMPORAL DIFFERENCE LEARNING


## Algorithm (contd.)

For each step in episode:

- Discretize state
- Select action acc to given policy
- Take action
- Get reward, new state
- Define value of state

$$V(S_t) = V(S_t) + \alpha(G_t - V(S_t))$$


# Algorithm (contd.)

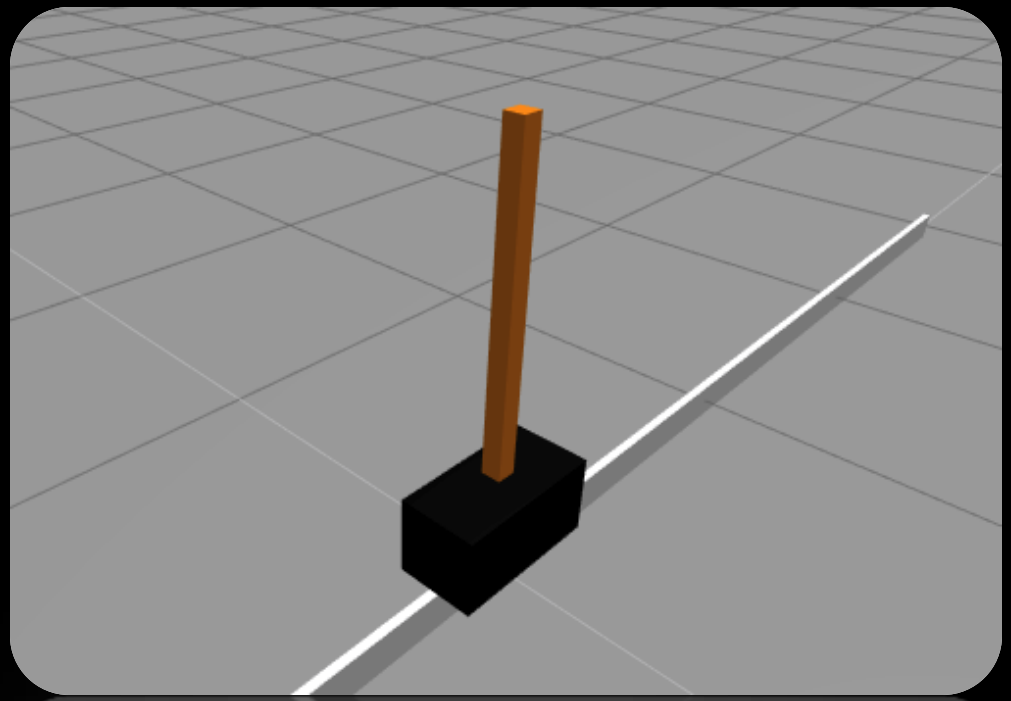
For each step in episode:

- Discretize state
- Select action acc to given policy
- Take action
- Get reward, new state
- Define value of state
- Update state to new state

# Robotics & AI Meetup #12

## Cart Pole Balance using Q-Learning

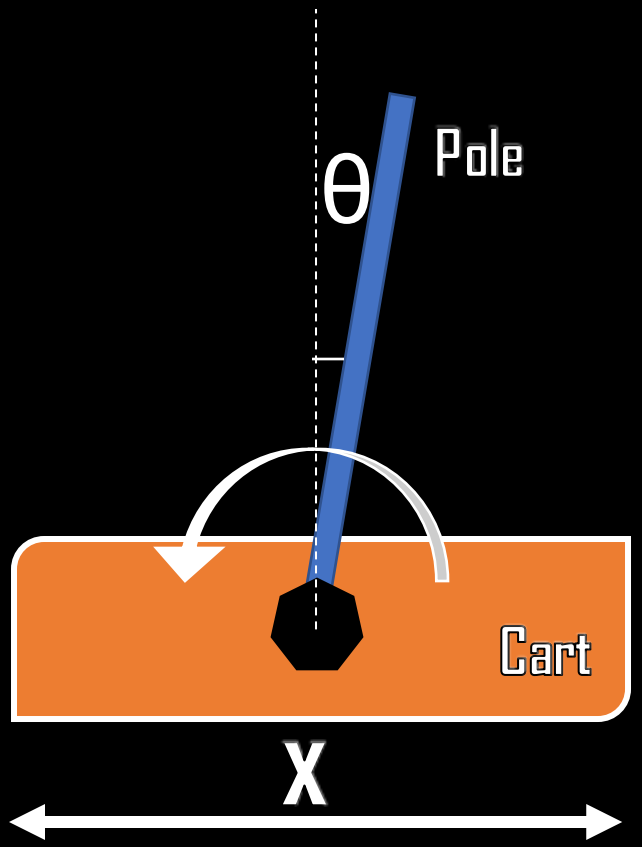
- Goal
- Problem Definition
- Brainstorming
- Project / Case Study Solution
- Algorithm + Code Walkthrough
- Concepts Covered
- Task for next week



BEGINNER FRIENDLY

THE BUILDING CULTURE

# Problem Definition




## State

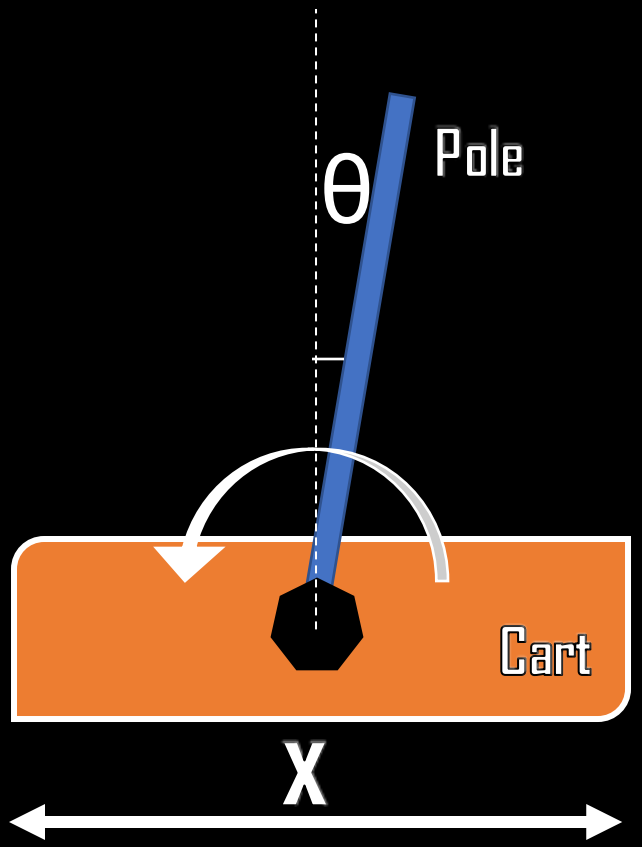
[x, x_dot, θ, θ_dot]

- x – Cart position
- x_dot – Cart velocity
- θ – Pole position
- θ_dot – Pole tip velocity

## State Space

- Range of allowable values for each state variable
- [-2.4, -inf, -41.8, -inf] to [2.4, inf, 41.8, inf]

# Problem Definition




## Action

[0, 1]

- 0 – Moving x left
- 1 – Moving x right

## Reward

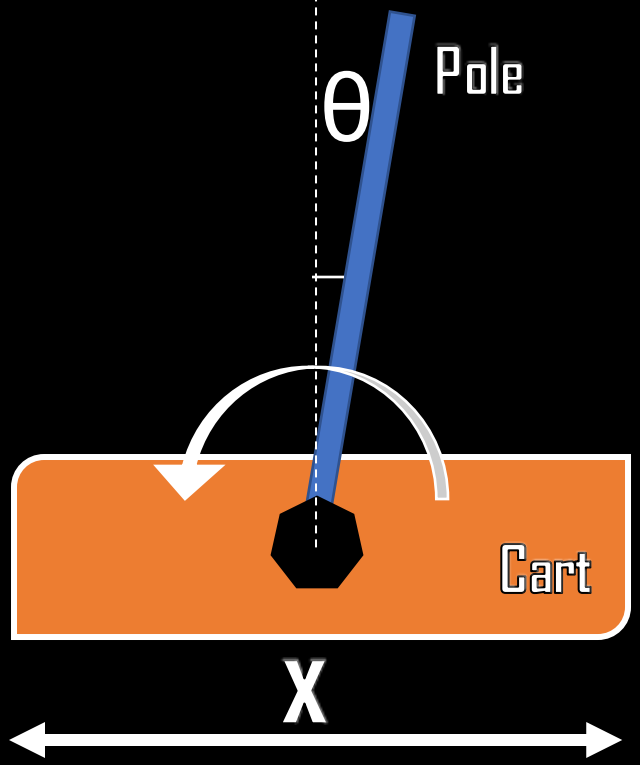
R(state, action, next_state) = +1

- Reward of +1 every step when it keeps |θ| <= 12 deg (Pole upright)

## Goal reached

When running average value of rewards > 195

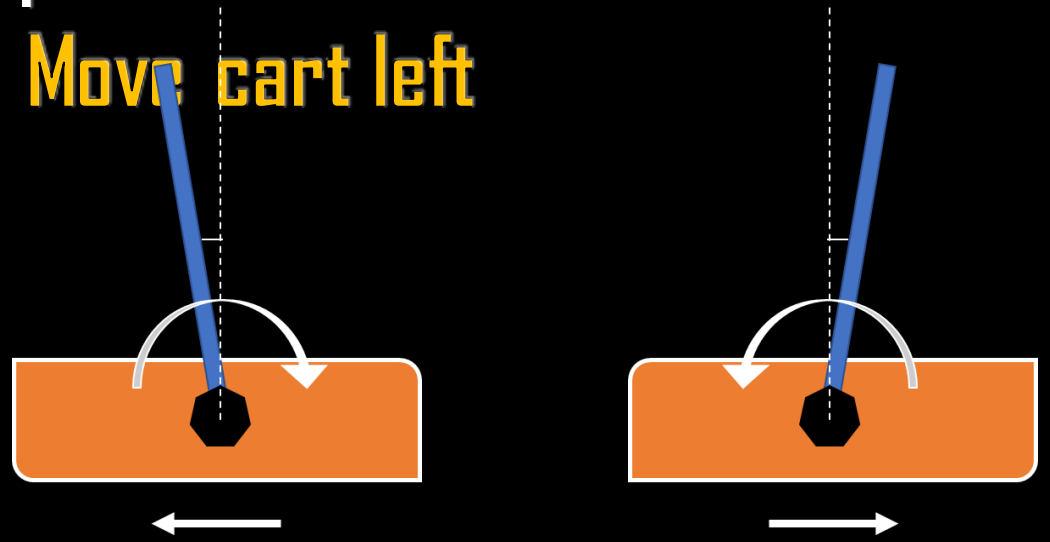
# Problem Definition




## Policy

- Epsilon Greedy Policy
- Random action in the beginning
- Action with max Value later
- If pole is left of the center

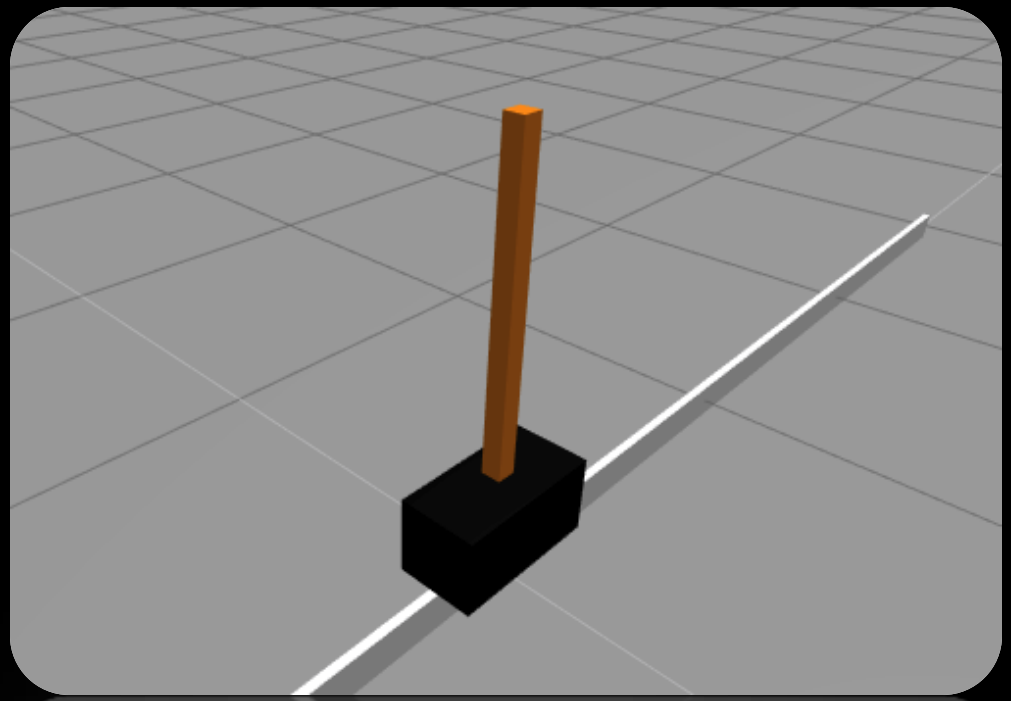
  Move cart left



# Robotics & AI Meetup #12

## Cart Pole Balance using Q-Learning

- Goal
- Problem Definition
- Brainstorming
- Project / Case Study Solution
- Algorithm + Code Walkthrough
- Concepts Covered
- Task for next week



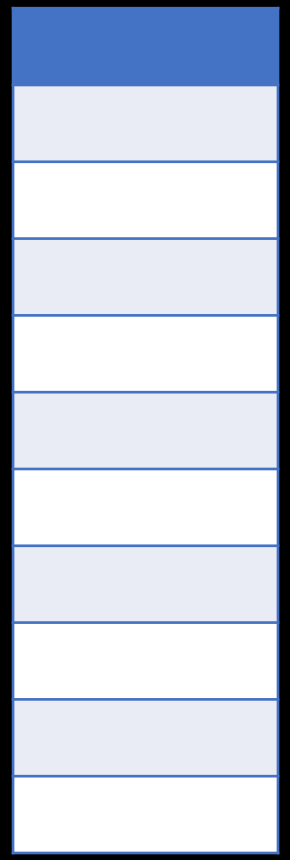
BEGINNER FRIENDLY

THE BUILDING CULTURE

CONCEPT : PREDICTION AND CONTROL ⚠

# Algorithm – Temporal Difference Control

- Define the policy
- Initialize arbitrary Q, terminal state 0

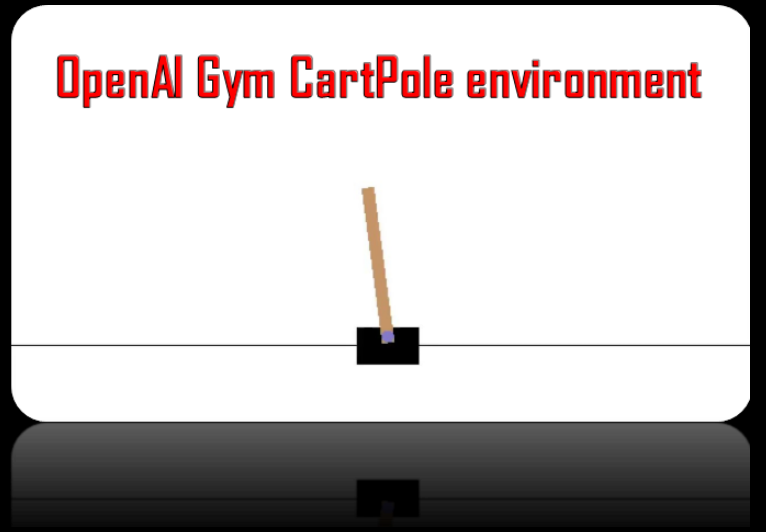


- If pole is left of the center
  Move cart left
- If pole is right of the center
  Move cart right

# CONCEPT : PREDICTION AND CONTROL ⚠

## Algorithm – Temporal Difference Control

- Initialize the policy
- Initialize arbitrary Q, terminal state 0
- For N no. of episodes :
- initialize environment



CONCEPT : PREDICTION AND CONTROL ⚠

# Algorithm – Temporal Difference Control

- Initialize the policy
- Initialize arbitrary Q, terminal state 0
- For N no. of episodes :
    - initialize environment
    - For each step in episode:
        - Discretize state
        - Select action acc to epsilon greedy policy
        - Take action

CONCEPT : PREDICTION AND CONTROL ⚠

# Algorithm – Temporal Difference Control

- Initialize the policy
- Initialize arbitrary Q, terminal state 0
- For N no. of episodes :
    - initialize environment
    - For each step in episode:
        - Discretize state
        - Select action acc to epsilon greedy policy
        - Take action

Epsilon starts at 1, goes down to 0.01 halfway through

# CONCEPT : Q-LEARNING ⚠

# Algorithm (contd.)

THE BUILDING CULTURE

For each step in episode:

- Discretize state
- Select action acc to given policy
- Take action
- Get reward, new state
- Q(St,At) = Q(St, At) + alpha[Rt+1 + gamma max Q(St+1,a) – Q(St, At)]
-

# CONCEPT : Q-LEARNING ⚠

# Algorithm (contd.)

THE BUILDING CULTURE

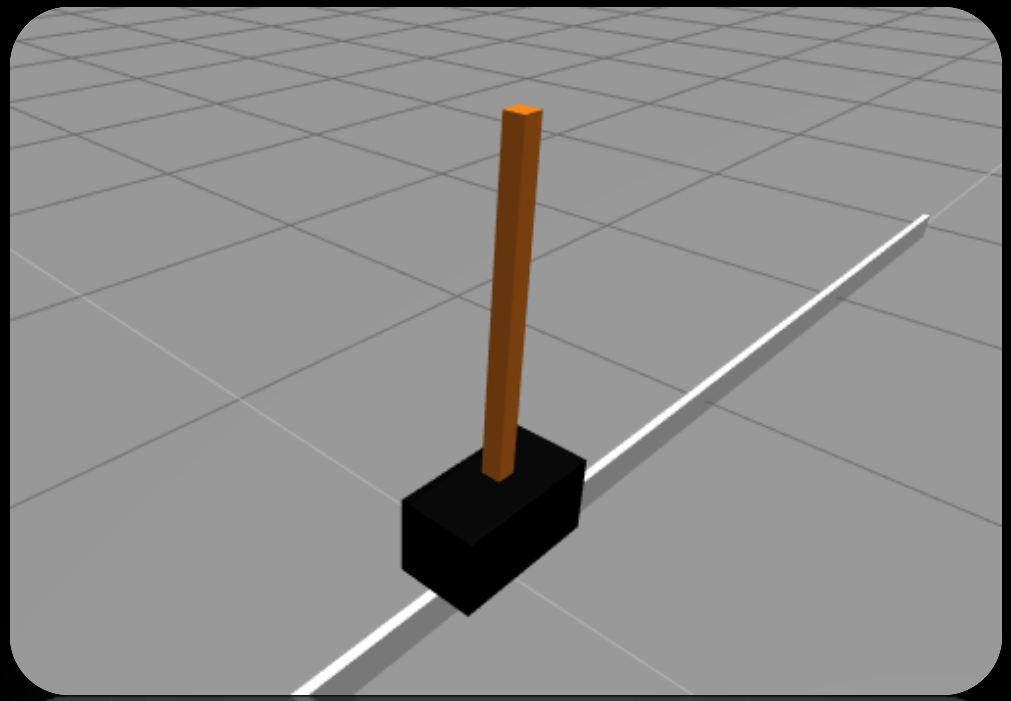
For each step in episode:

- Select action acc to given policy
- Take action
- Get reward, new state
- $Q(St,At) = Q(St, At) + alpha[Rt+1 + gamma \max Q(St+1,a) - Q(St, At)]$
- Update state to new state
-

# Robotics & AI Meetup #10

## Cart Pole Balance using Reinforcement Learning

- Goal
- Problem Definition
- Brainstorming
- Project / Case Study Solution
- Algorithm + Code Walkthrough
- Concepts Covered
- Task for next week



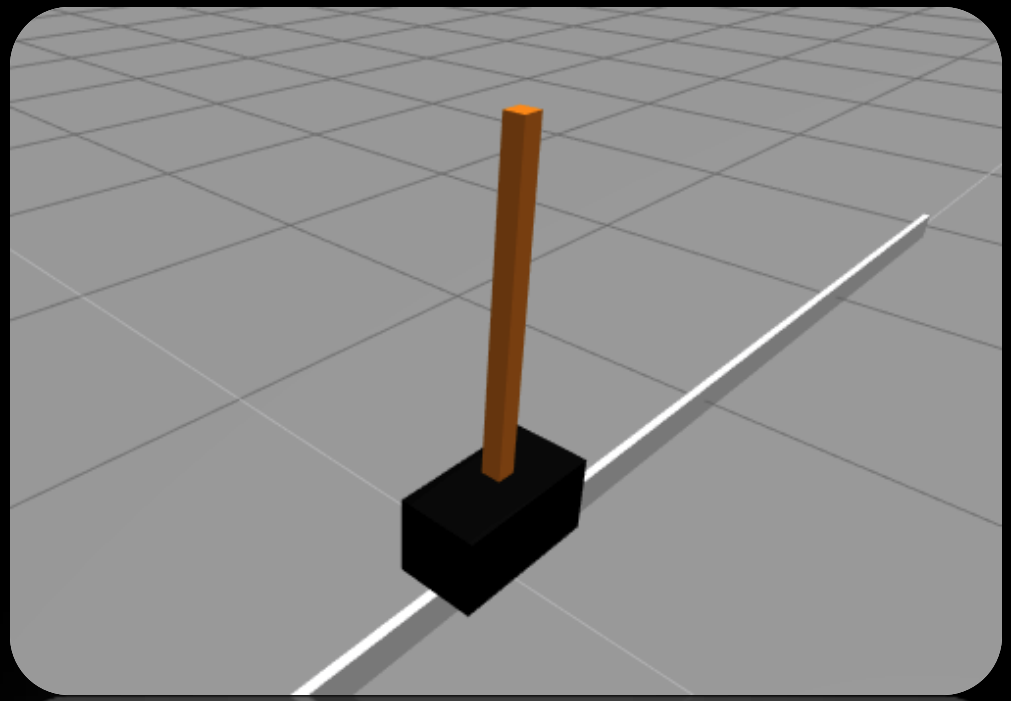
BEGINNER FRIENDLY

THE BUILDING CULTURE

# Robotics & AI Meetup #10

## Cart Pole Balance using Reinforcement Learning

- Goal
- Problem Definition
- Brainstorming
- Project / Case Study Solution
- Algorithm + Code Walkthrough
- Concepts Covered
- Task for next week



BEGINNER FRIENDLY

THE BUILDING CULTURE

# Meetup #12 – RL Concepts


RL PROBLEM

CONT. STATE SPACES

EXPECTED RETURN

TEMPORAL DIFF

Q - LEARNING

BELLMAN EQUATION

ACTION VALUE FUNCTION

STATE VALUE FUNCTION

MONTE CARLO PREDICTION

MODEL BASED RL

MODEL FREE RL

MARKOV DECISION PROCESS

PREDICTION AND CONTROL

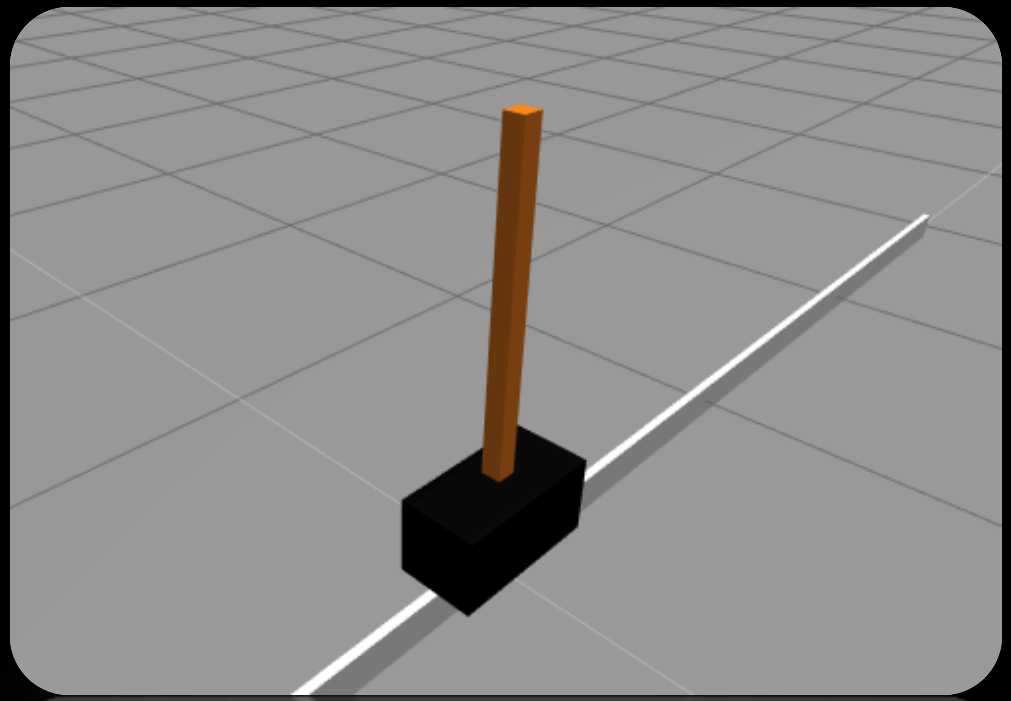
MONTE CARLO CONTROL

EPSILON GREEDY POLICY

# Robotics & AI Meetup #10

## Cart Pole Balance using Reinforcement Learning

- Goal
- Problem Definition
- Brainstorming
- Project / Case Study Solution
- Algorithm + Code Walkthrough
- Concepts Covered
- Task



BEGINNER FRIENDLY

THE BUILDING CULTURE

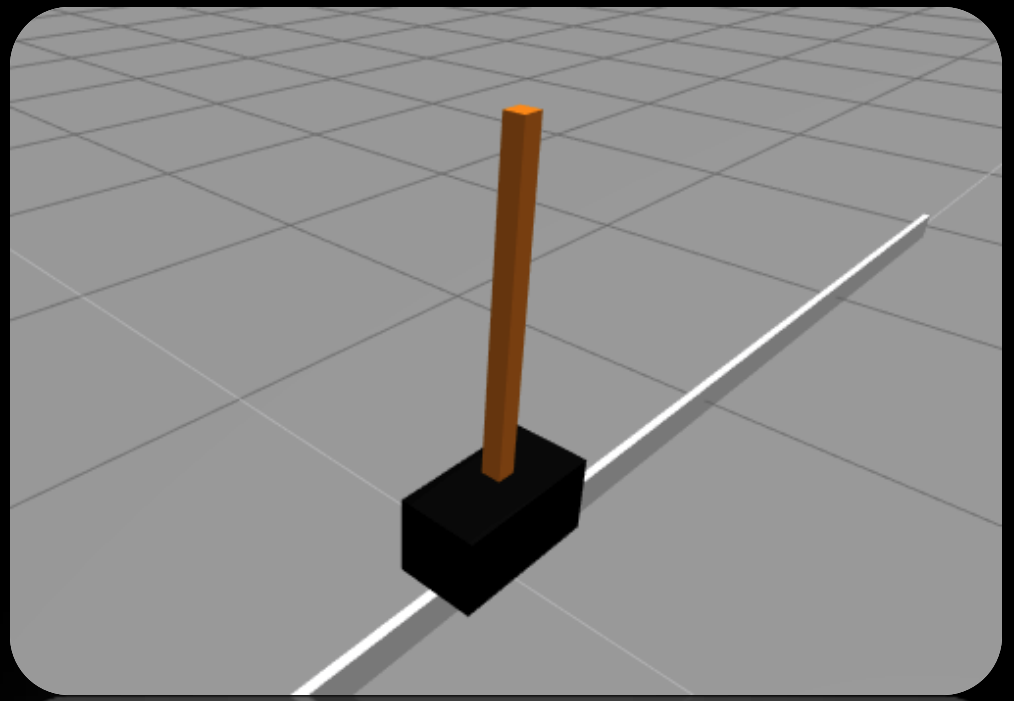
# TASK (OPTIONAL)


- Codebase shared on the Discord server
- Play with the code
- Write your own implementation
- Show progress / discuss in next meetup

# Robotics & AI Meetup #13

## LIVE Coding - Cart Pole Balance using Q-Learning

THE BUILDING CULTURE



Code together from scratch

Code shared beforehand

BEGINNER FRIENDLY

# REFERENCES


- https://spinningup.openai.com/en/latest/spinningup/rl_intro.htm
- Deep Lizard Youtube channel
- Machine Learning with Phil Youtube channel
- Steve Brunton Youtube channel
- David Silver Lectures

YOU CAN SWITCH TOPICS OR EVEN DO LATER IF YOU ARE UNABLE TO PREPARE

# MEETUP #10 : 06/03/2021

THE BUILDING CULTURE

# AGENDA

19:00 Introductory Presentation

19:05 Project Presentation

19:45 QnA

20:00 Feedback for Presentation

20:05 Feedback for Project Process

20:10 Project Accountability Share (if any)

20:15 Introductions and casual networking

20:30 Announcement

# MEETUP #10 : 06/03/2021

THE BUILDING CULTURE

## AGENDA

19:00 Introductory Presentation

19:05 Project Presentation

19:45 QnA

20:00 Feedback for Presentation

20:05 Feedback for Project Process

20:10 Project Accountability Share (if any)

20:15 Introductions and casual networking

20:30 Announcement

# MEETUP #10 : 06/03/2021

THE BUILDING CULTURE

## AGENDA

19:00 Introductory Presentation

19:05 Project Presentation

19:45 QnA

20:00 Feedback for Presentation

20:05 Feedback for Project Process

20:10 Project Accountability Share (if any)

20:15 Introductions and casual networking

20:30 Announcement

# MEETUP #10 : 06/03/2021

THE BUILDING CULTURE

## AGENDA

19:00 Introductory Presentation

19:05 Project Presentation

19:45 QnA

20:00 Feedback for Presentation

20:05 Feedback for Project Process

20:10 Project Accountability Share (if any)

20:15 Introductions and casual networking

20:30 Announcement

# MEETUP #10 : 06/03/2021

THE BUILDING CULTURE

## AGENDA

19:00 Introductory Presentation

19:05 Project Presentation

19:45 QnA

20:00 Feedback for Presentation

20:05 Feedback for Project Process

20:10 Project Accountability Share (if any)

20:15 Introductions and casual networking

20:30 Announcement

# MEETUP #10 : 06/03/2021

THE BUILDING CULTURE

## AGENDA

19:00 Introductory Presentation

19:05 Project Presentation

19:45 QnA

20:00 Feedback for Presentation

20:05 Feedback for Project Process

20:10 Project Accountability Share (if any)

20:15 Introductions and casual networking

20:30 Announcement

# MEETUP #10 : 06/03/2021

THE BUILDING CULTURE

## AGENDA

19:00 Introductory Presentation

19:05 Project Presentation

19:45 QnA

20:00 Feedback for Presentation

20:05 Feedback for Project Process

20:10 Project Accountability Share (if any)

20:15 Introductions and casual networking

20:30 Announcement

# PLEASE FILL THE ROLES

THE BUILDING CULTURE

- Roles filled - News-feed updater
- Open roles -
- Project presenter (*#meetup-volunteer* in Events)
- Research Paper presenter
  (*#paper-reading-volunteer* in Events)

**VOLUNTEER-DRIVEN COMMUNITY !**

MEETINGS WILL BE RECORDED AND SHARED ON YOUTUBE

EVERYONE CAN PARTICIPATE REGARDLESS OF EXPERIENCE OR AGE GROUP


# THANK YOU