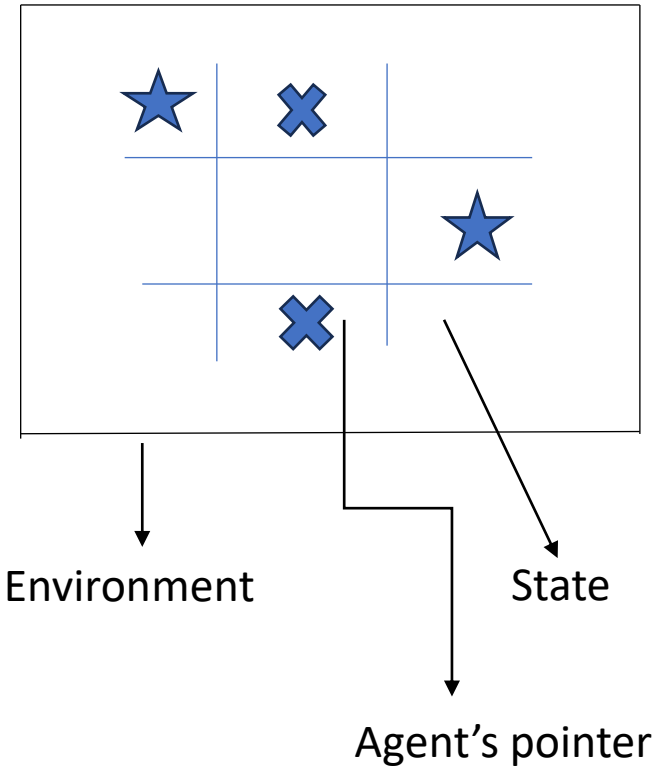
Comparison of Reinforcement Learning Algorithms for Continuous Problem

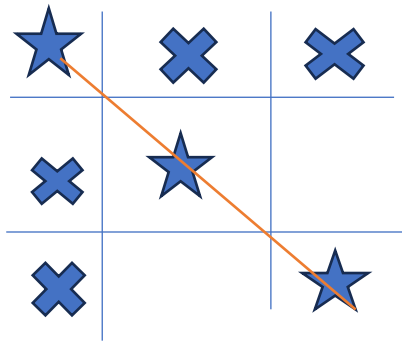
Swati Kar, Mahesh Banavar, Soumyabrata Dey

Outline

- Introduction
- Problem definition
- Research objective
- Algorithm Description
- Result Analysis
- Discussion

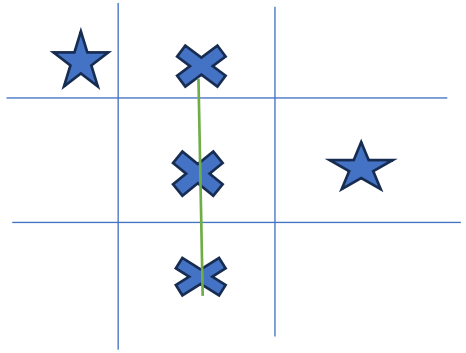
What is Reinforcement Learning?





Lose!

Reward: -1

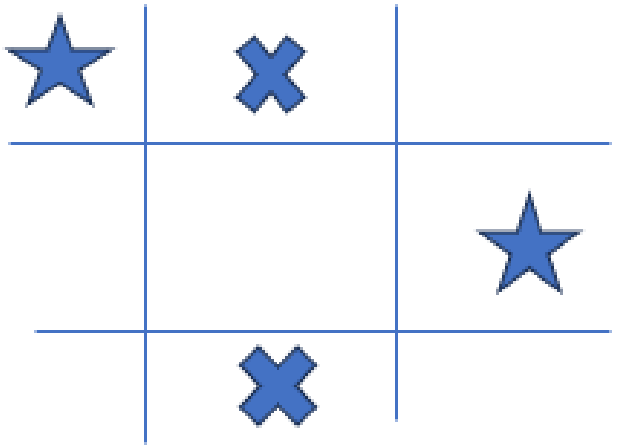


Win!

Reward: +1

Why Reinforcement Learning?

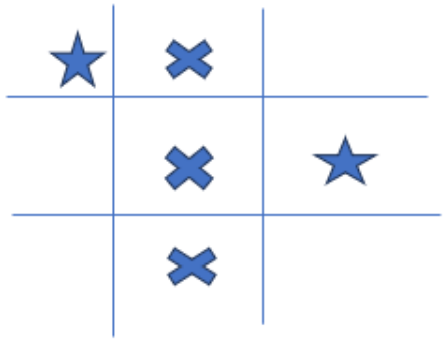
- Well suited when the environment is dynamic
- Algorithm can easily adapt environment
- No prior knowledge is required, receives feedback from the environment in the form of reward
- Due to diverse adaptability, Reinforcement Learning is applied in
  - Robotics
  - Finance
  - Healthcare etc.



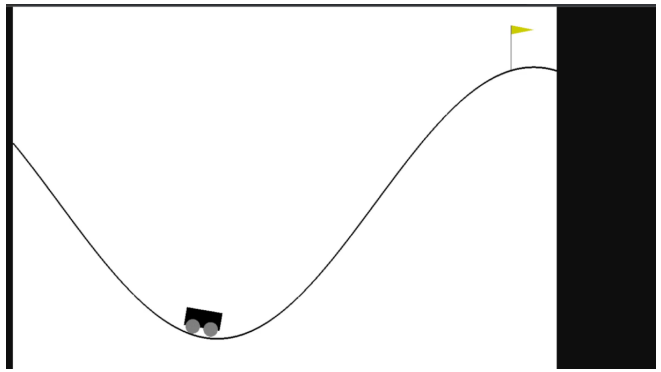
Research Objective

In Reinforcement Learning, two types of environment are used:

Discrete



Continuous

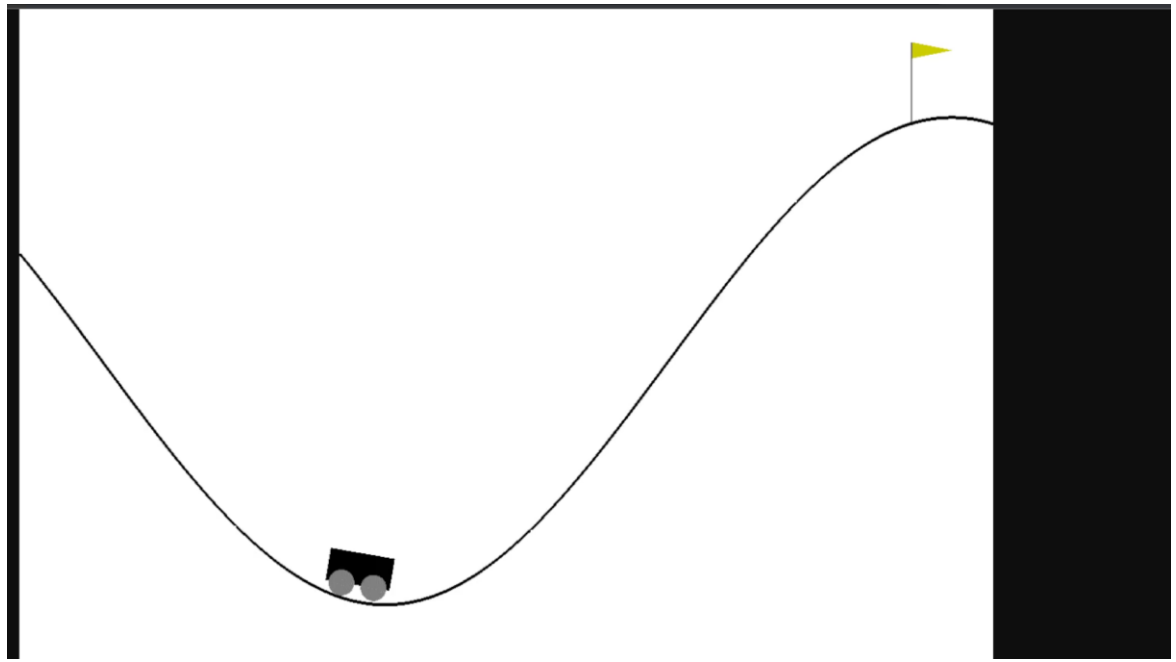


Our research objective is comparison among three prominent RL algorithms for Mountain car continuous problem:

- Q-Learning
- Deep Q Learning (DQN)
- Actor-Critic

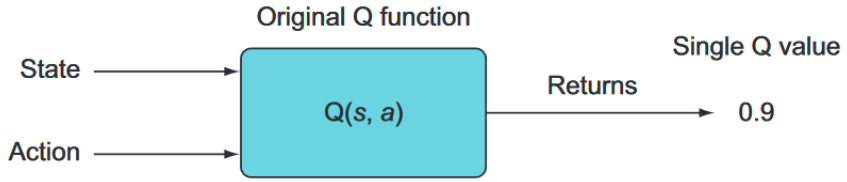
Problem Definition

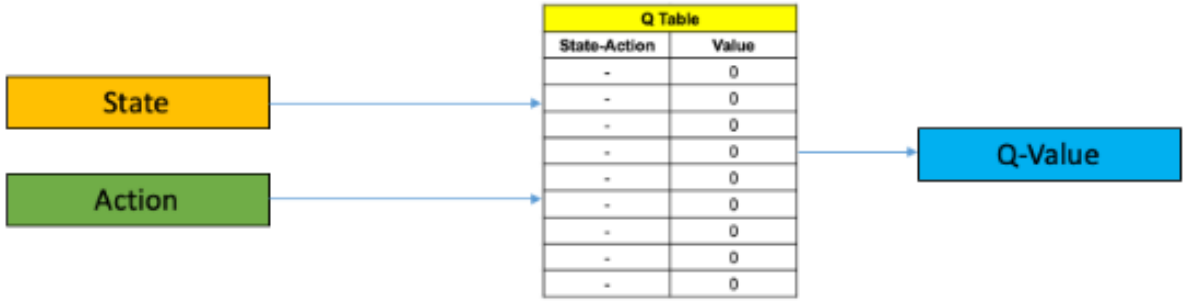
- We used OpenAI Gym's Mountain Car environment
- Agent must learn to navigate a car up a steep hill
- State space (position:[0.6 -1.2] and velocity:[0.07 -0.07]) and action space (action: [-1.0 1.0]) of this environment are continuous, that makes it challenging
- If it reaches to the goal position, +100 reward will be provided
- Otherwise, it will get negative reward



Q-Learning Algorithm

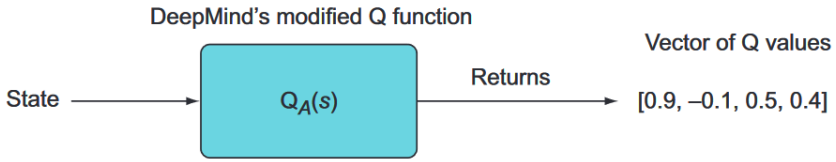
- Value based algorithm that learns Q-table
- Maps state-action pairs to expected rewards





Deep Q Learning Algorithm

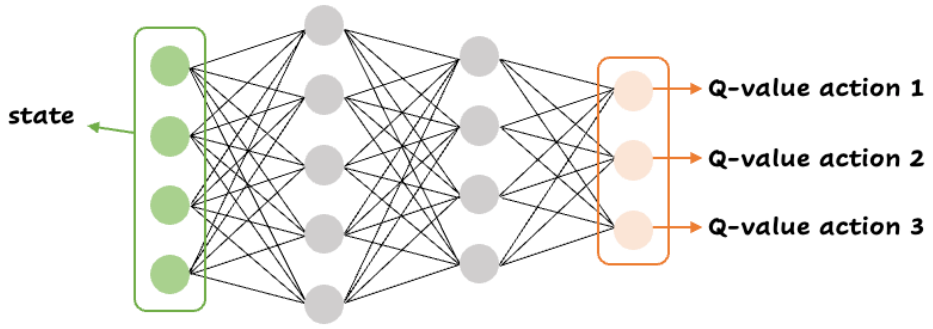
- Uses a deep neural network to represent the Q-table
- Takes only states as input and produce output for every action



Tabular Q-learning

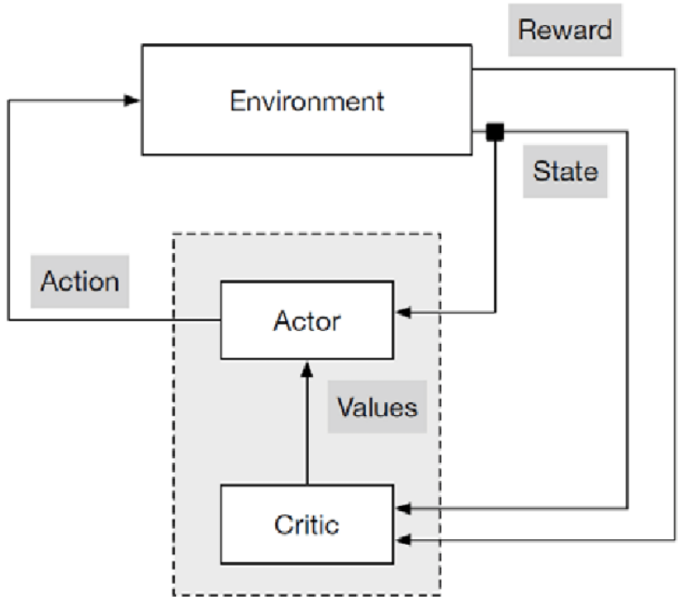
	Actions			
States	1	2	...	n
0	Q(0,1)	Q(0,2)	...	Q(0,n)
1	Q(1,1)	Q(1,2)	...	Q(1,n)
...	...	...	...	...
m	Q(m,1)	Q(m,2)	...	Q(m,n)

Deep Q-learning



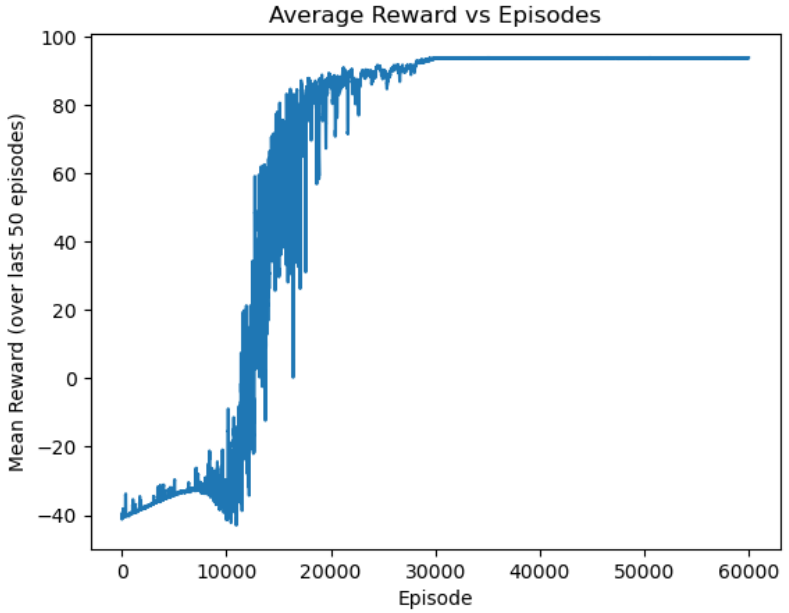
Actor-Critic Algorithm

- Learns a policy and a value function simultaneously
- The actor learns a policy that maps states to actions, and the critic learns a value function that estimates the expected reward of a state.



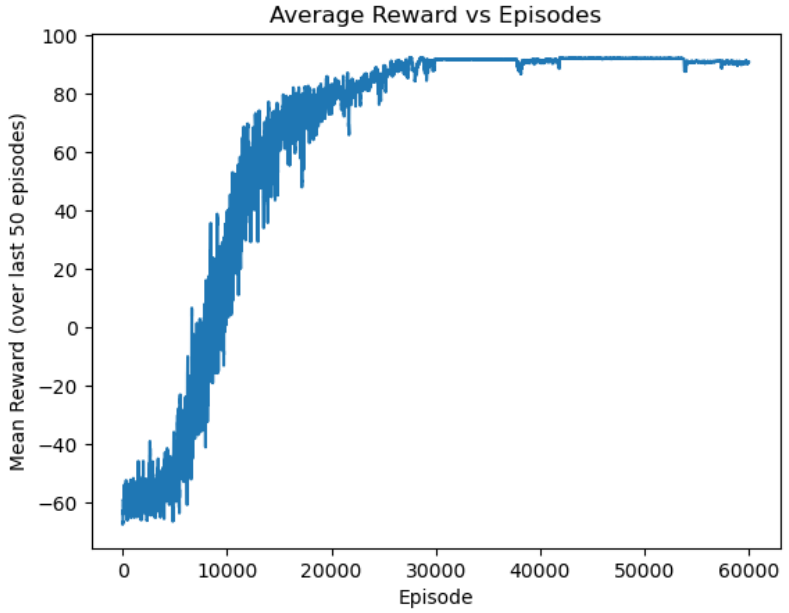
Results (Q-Learning Algorithm)

- Environment Setup:
  - State space
    - Divided position into 12 equal chunks
    - Divided velocity into 20 equal chunks
  - Divided action space into 3 equal chunks



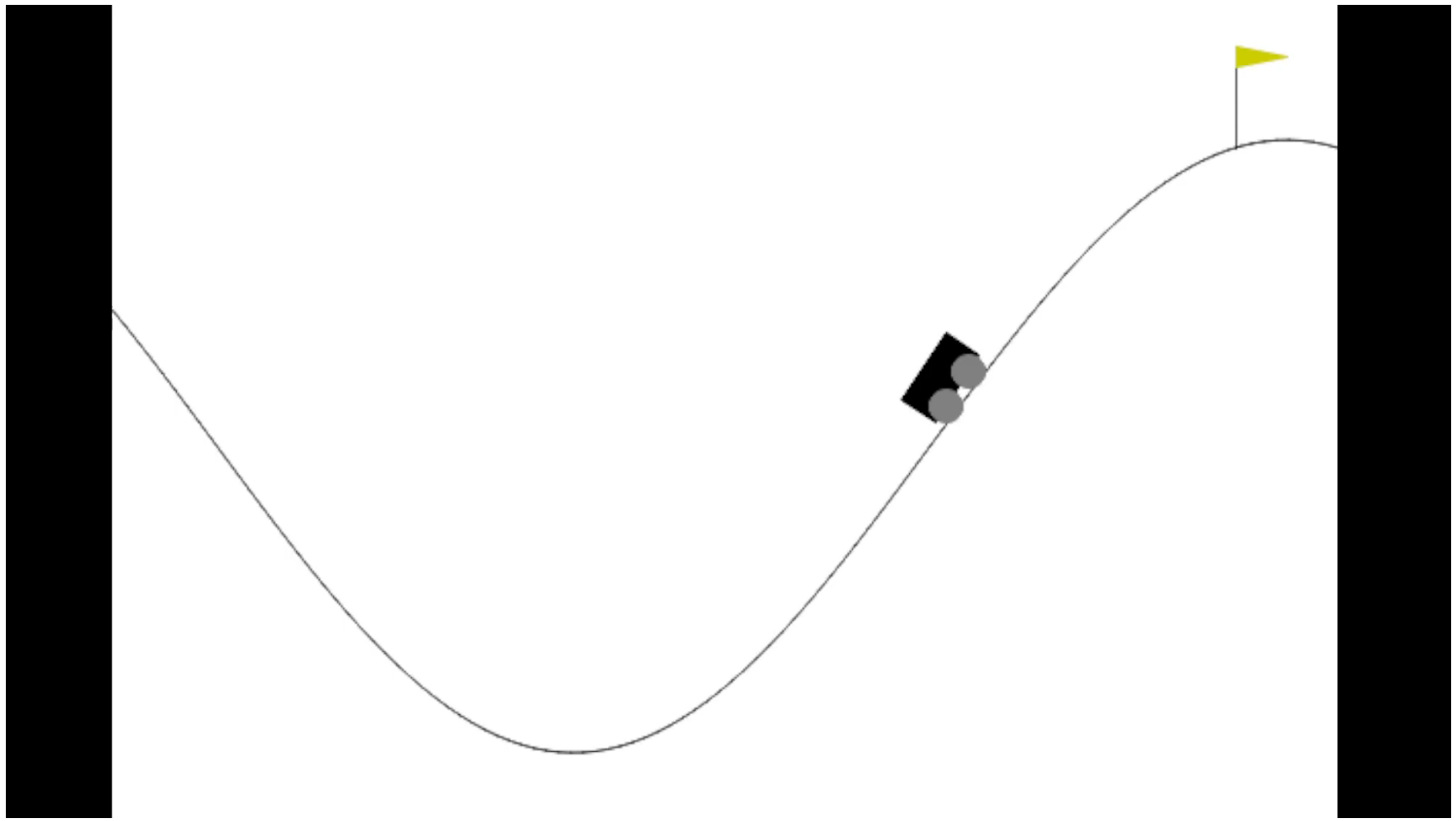
Converged at 21141 episode

- Environment Setup:
  - State space
    - Divided position into 16 equal chunks
    - Divided velocity into 20 equal chunks
  - Divided action space into 9 equal chunks



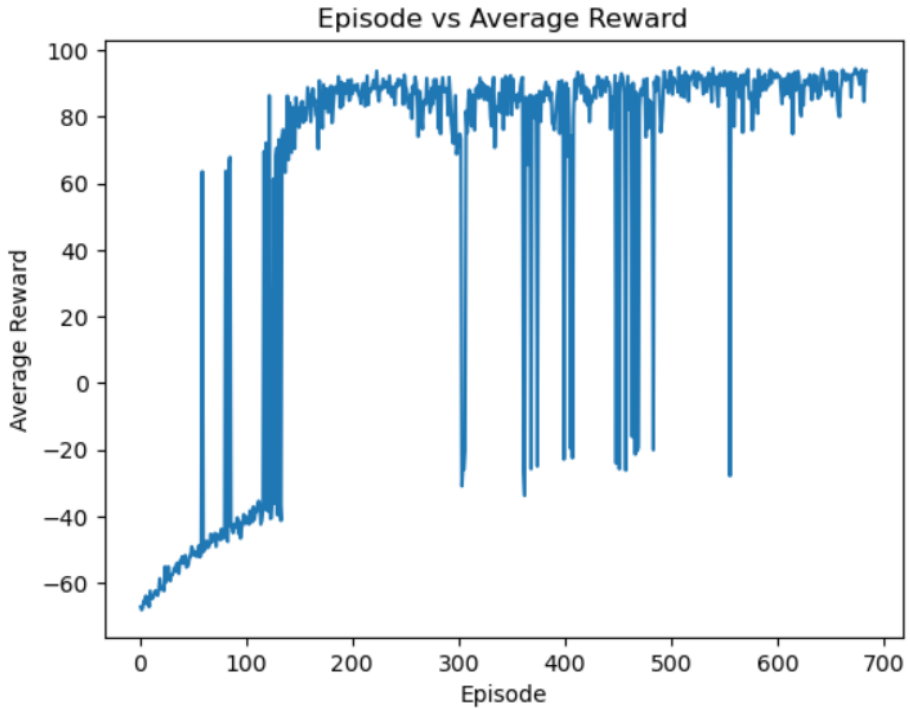
Converged at 23806 episode

Results (Q-Learning Algorithm)



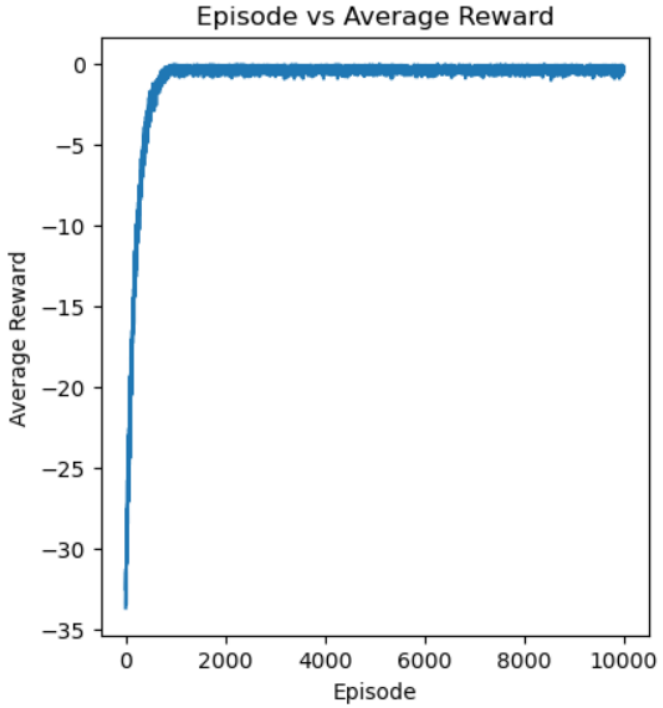
Results (Deep Q learning)

- Environment Setup:
  - State space is continuous
  - Divided action space into 3 equal chunks



Converged at 682 episode

- Environment Setup:
  - State space is continuous
  - Action space is continuous

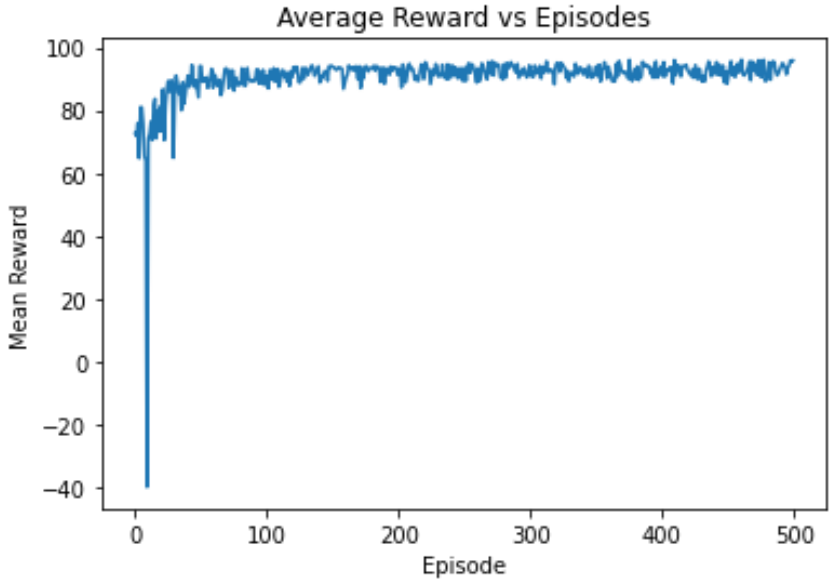


Not Converged (expected)

Results (Actor-Critic Algorithm)

Environment Setup:

- Continuous action space
- Continuous state space



Converged at episode 56

Comparison

Q-Learning Algorithm	Deep Q learning Algorithm	Actor Critic Algorithm
Not suitable for • Continuous state space • Continuous action space	Suitable for • Continuous state space • Not suitable for continuous action space	Suitable for both • Continuous state space • Continuous action space
Converged approximately 21K episodes	Converged at 682 episode for discrete action space	Converged at 56 episode

Conclusion

- Among three algorithms, Actor-Critic performs significantly better than other two algorithms when environment is continuous
- In future, we will explore more continuous problems

Thank You! Any Questions?

References

- https://www.google.com/search?rlz=1C1VDKB_enUS1033US1033&sxsrf=AB5stBgT956SvHcRCOECbaey_QqEIGGKqg:1690134820210&q=image+of+dqn&tbm=isch&sa=X&ved=2ahUKEwjV4dLssqWAAxXug4kEHVy0CnwQ0pQJegQIDBAB&biw=1600&bih=781&dpr=1#imgrc=5d7kyHEoKgRbOM&vwlns=WyIwQ0JFUWg2Y0dhaGNLRXdpUTh1bnVzcVdBQXhVQUFBQUFIUUFBQUFBUUJBIl0=&lns=W251bGwsbnVsbCxudWxsLG51bGwsbnVsbCxudWxsLG51bGwsIkVrY0tKREJrTlRneU5HRm1MV05rWTJVdE5EUXpNQzFoTVdZNUxUbGlNelF6T0RreFlUVTBOUklmVVRrNVZISkxVV0ZIUW1kYVNVVXhUR3gwZUVKVlRIQnZhRVJGTFcxQ1p3PT0iLG51bGwsbnVsbCxudWxsLDEsbnVsbCxbbnVsbCxudWxsLFswLDUyNDEyLDQ3NTg4LDEwMDAwMF0sbnVsbCw1XV0=
- https://www.google.com/search?q=image+of+q+learning&tbm=isch&ved=2ahUKEwjd_Z3usqWAAxW-J1kFHarVAVYQ2-cCegQIABAA&oq=image+of+q+learning&gs_lcp=CgNpbWcQAzoECCMQJzoFCAAQgAQ6CQgAEBgQgAQQCjoICAAQgAQQsQM6BggAEAgQHjoHCAAQGBCABFC1BliDKGDrKmgAcAB4AIAB2AGIAcMMkgEFOC41LjGYAQCgAQGqAQtnd3Mtd2l6LWltZ8ABAQ&sclient=img&ei=J2m9ZJ3OIL7P5NoPqquHsAU&bih=781&biw=1600&rlz=1C1VDKB_enUS1033US1033#imgrc=0shAK1RDkZZWNM