Artificial Intelligence Project

CARTPOLE BALANCING

Presented by

Anubhav Khanal
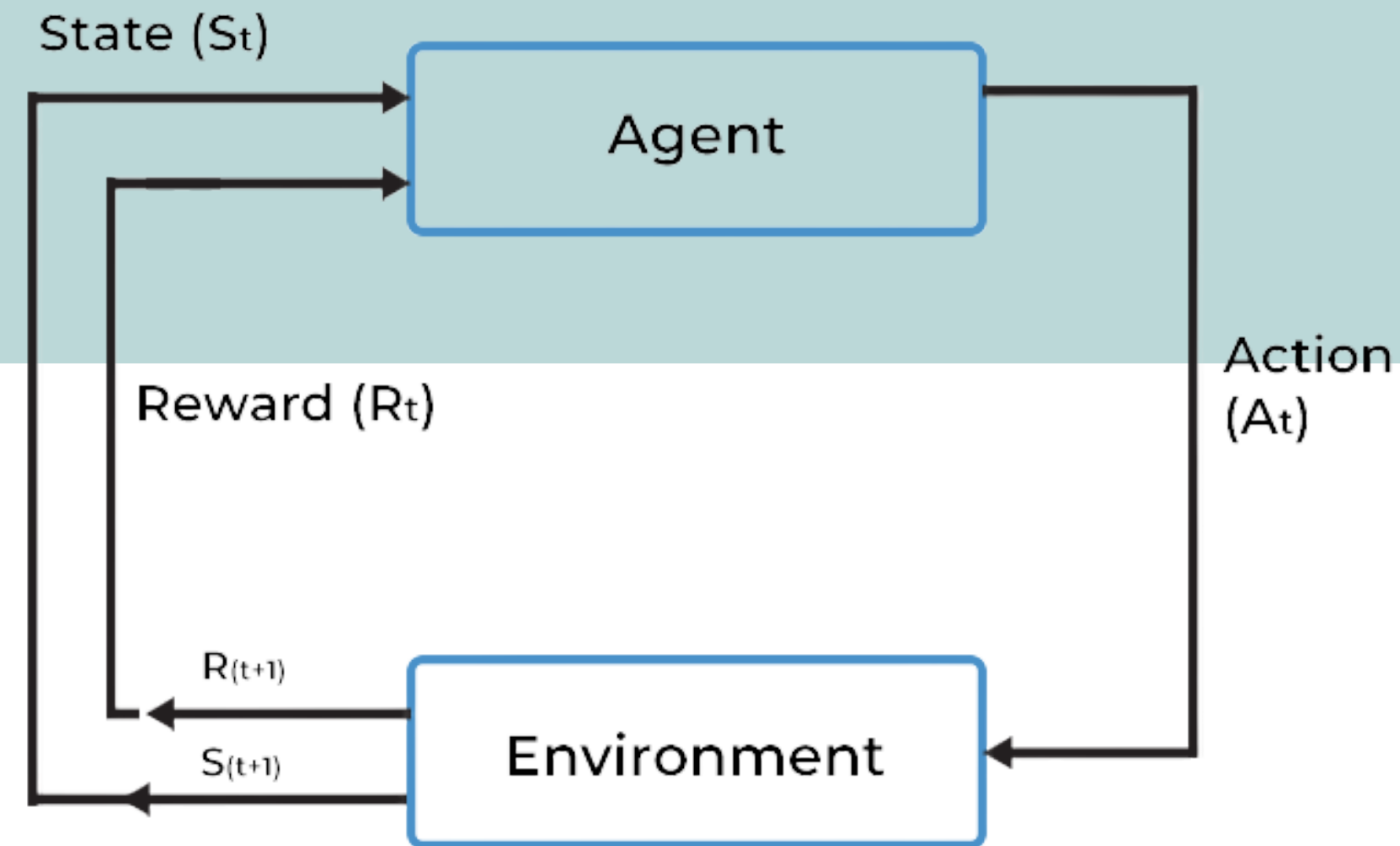Mahima Dhakal
Nadika Poudel
Pramish Paudel
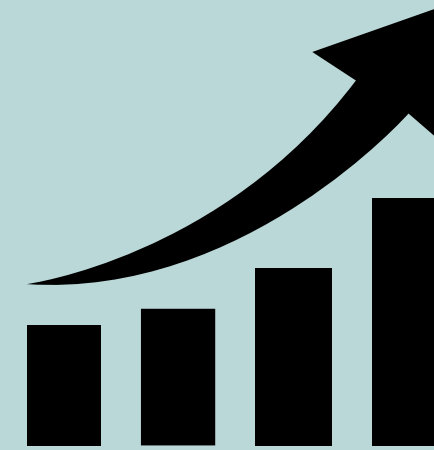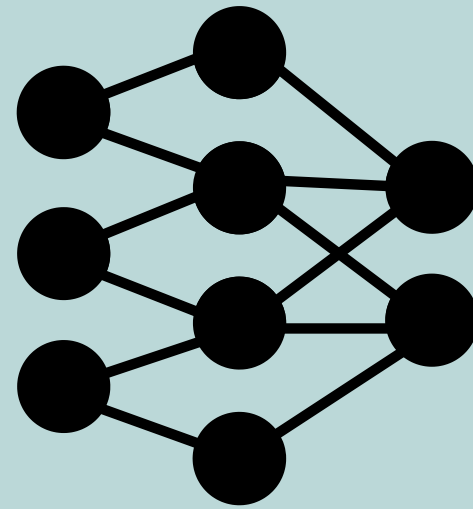
Reinforcement Learning

State (S_t)

Agent

Action (A_t)

Reward (R_t)

$R_{(t+1)}$

$S_{(t+1)}$

Environment

> A type of machine learning that involves an agent learning to make decisions by interacting with an environment where it takes actions in order to maximize a reward signal.
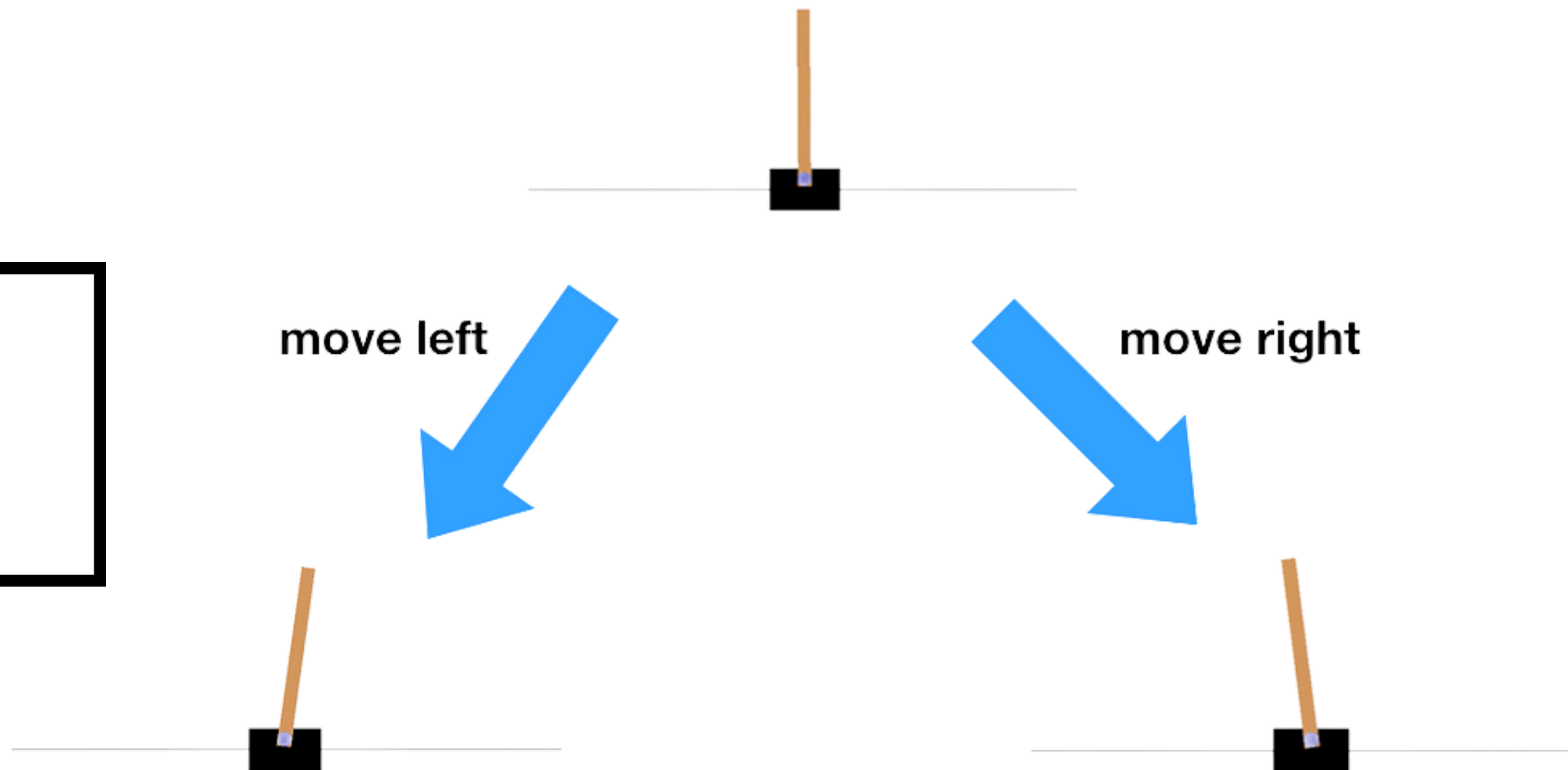
Introduction

What is DQN?

Neural Network used in Reinfrocement Learning

learn the best action to take in the given state

Used for estimating Q-value function

Cartpole Game

move left

move right

Experience Replay

$$S_t, A_t, R_{t+1}, S_{t+1}, P_t$$
$$S_{t+1}, A_{t+1}, R_{t+2}, S_{t+2}, P_{t+1}$$
$$S_{t+2}, A_{t+2}, R_{t+3}, S_{t+3}, P_{t+3}$$
$$S_{t+3}, A_{t+3}, R_{t+4}, S_{t+4}, P_{t+4}$$
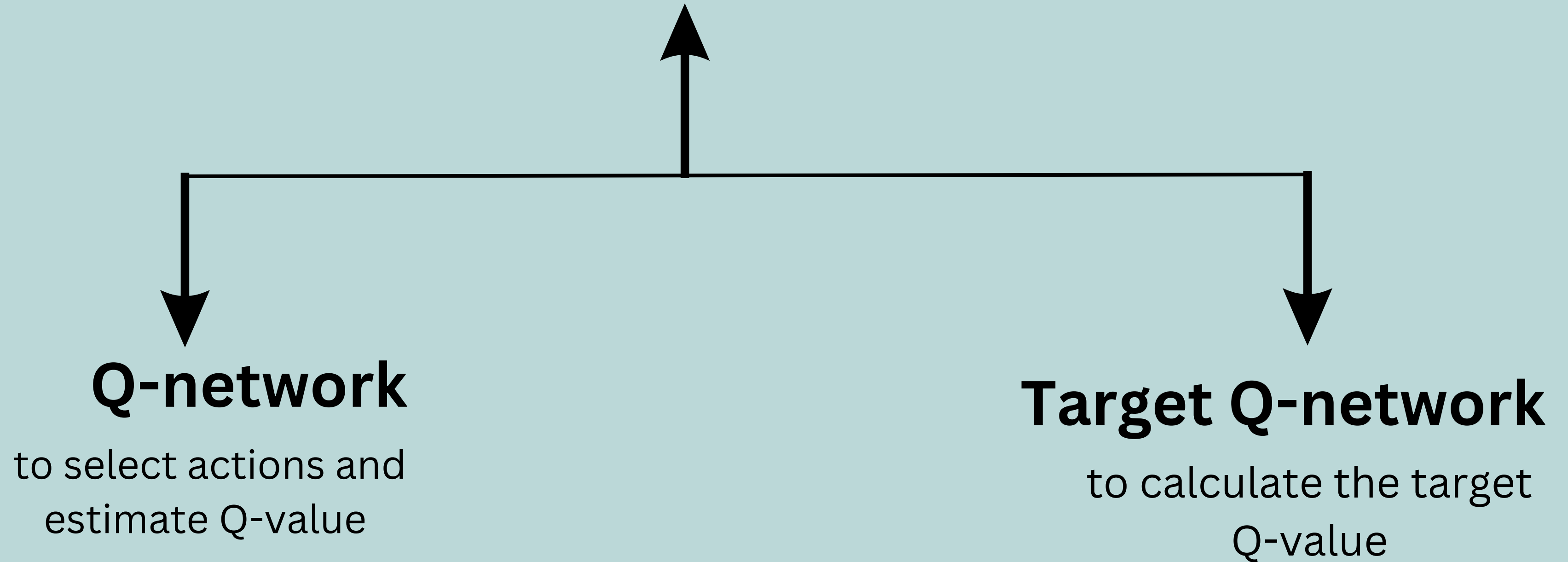...

Experience replay Buffer
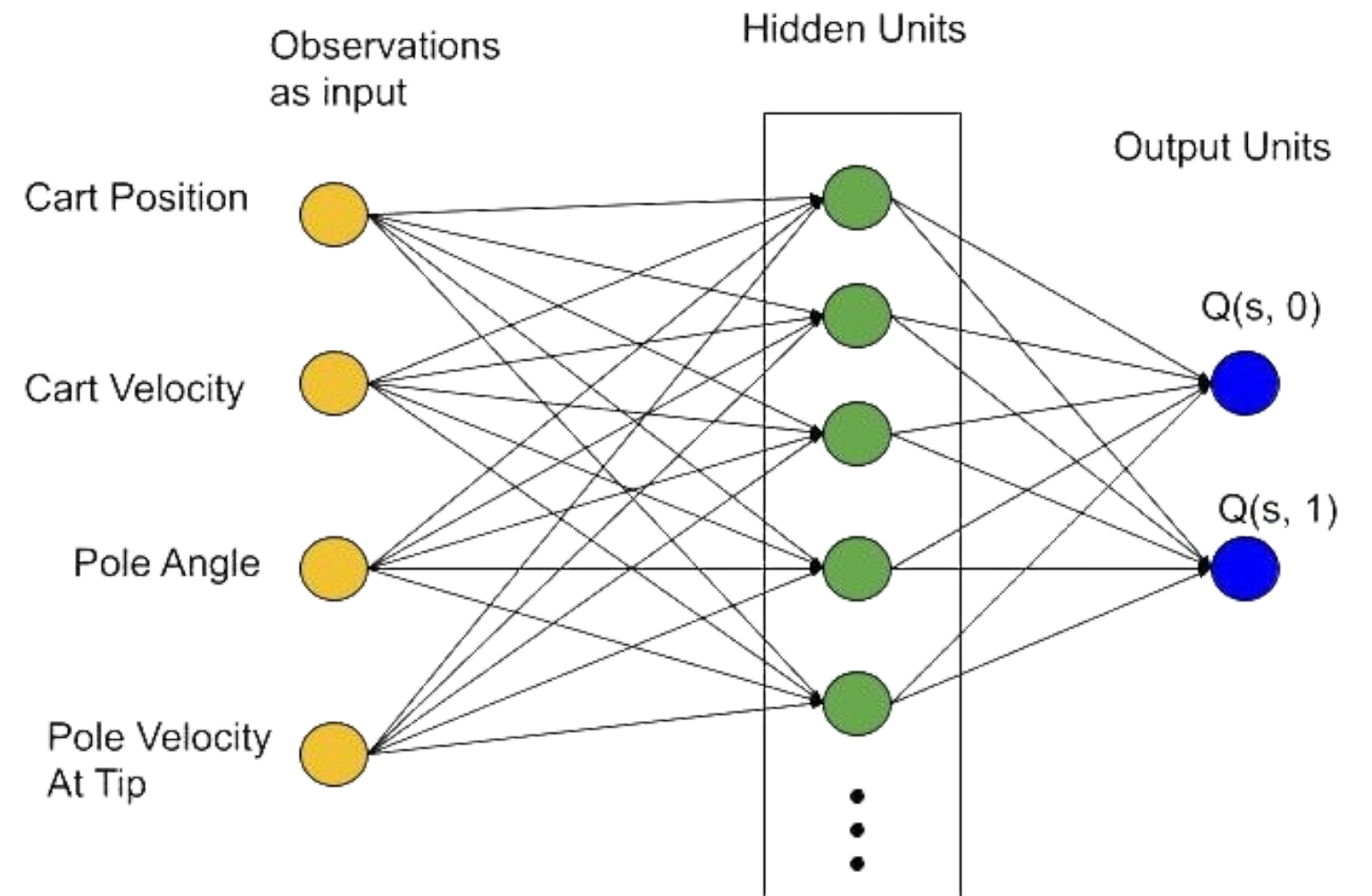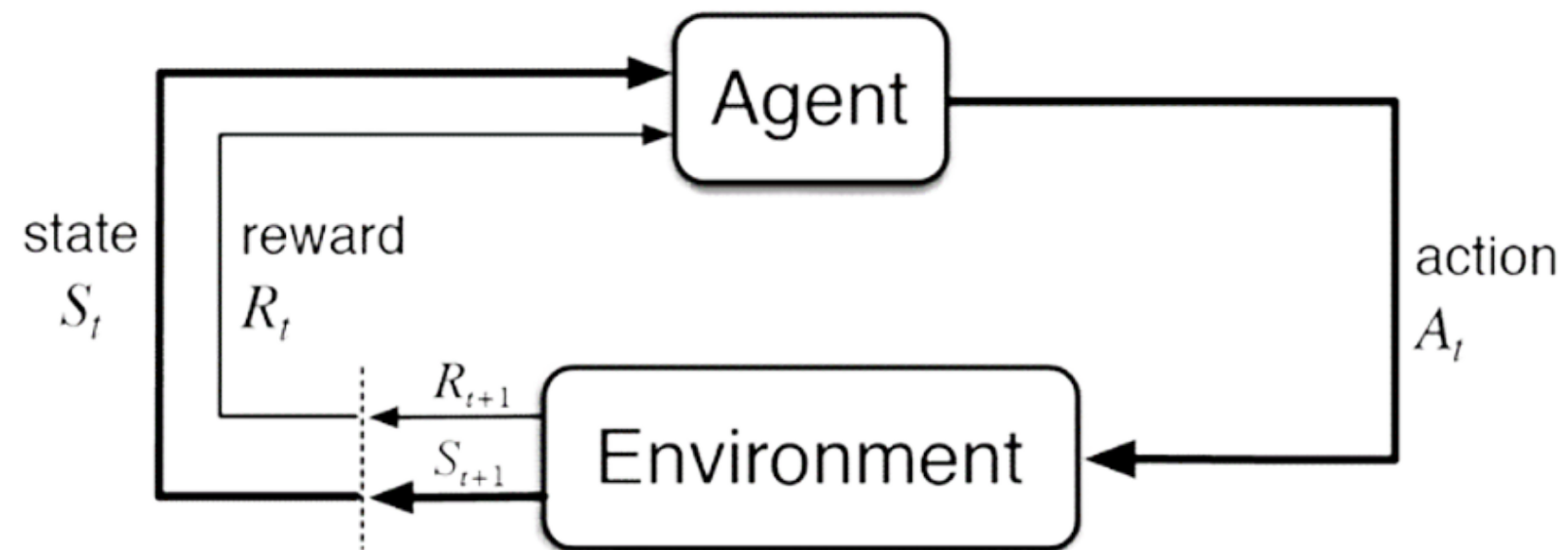
Sample

Batch of experiences

DQN Agent

- involves storing past experiences (states, actions, rewards, next states) in a replay buffer.

- agent randomly samples experiences from the buffer during training that breaks temporal correlations and prevents the agent from becoming too reliant on specific experiences or trajectories.

- By drawing from a diverse set of experiences, agents can learn to generalize better and perform better on new tasks.
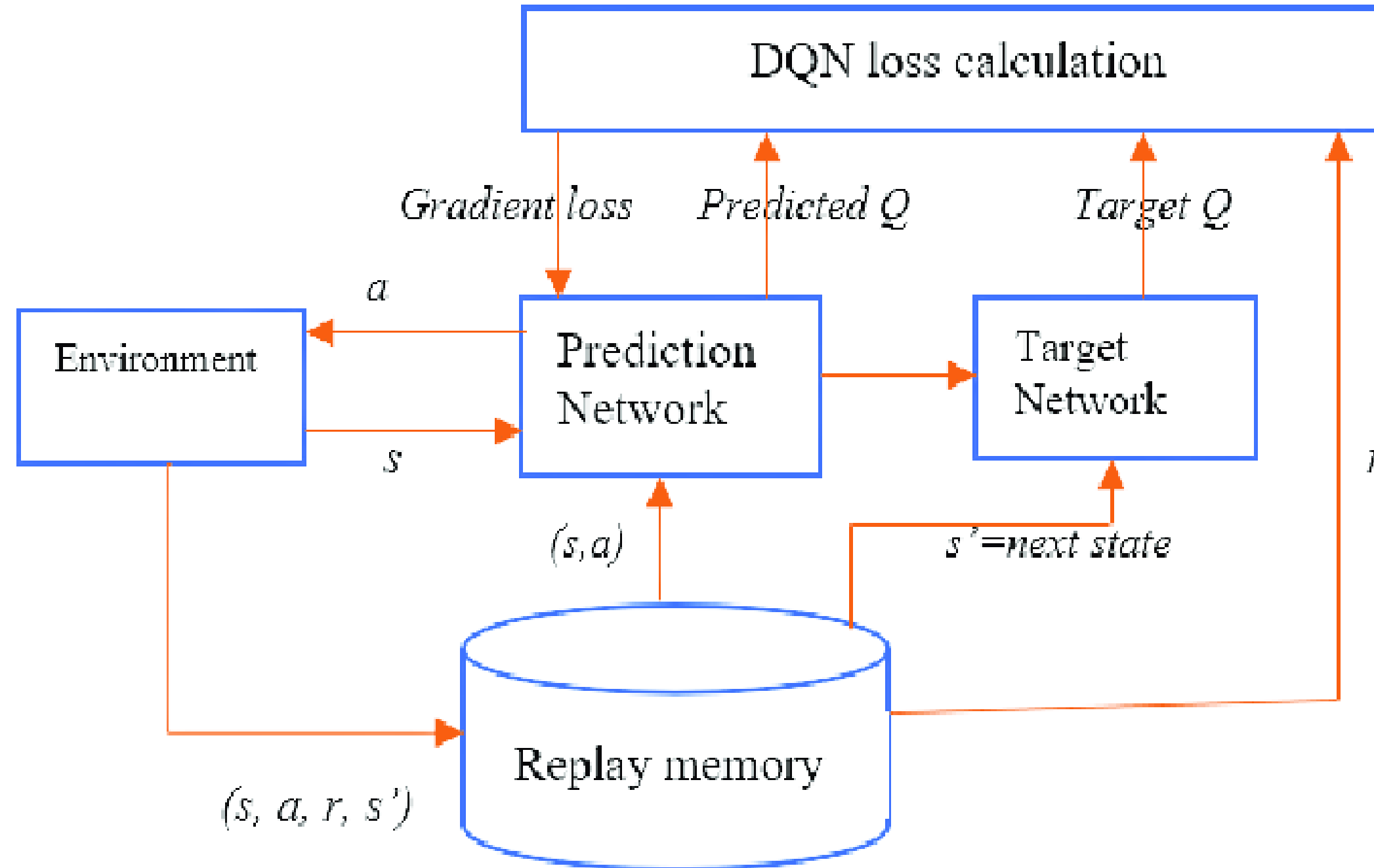
Neural Networks

Q-network

to select actions and
estimate Q-value

Target Q-network

to calculate the target
Q-value

Post Condition: Target network is updated in intervals

Neural Network Architecture

Data flow in a DQN

Challenges

- **Non-linearity:** Complex and nonlinear behavior of the cartpole system makes it difficult to model accurately.

- **High-dimensional state space:** Four continuous variables result in a high-dimensional state space

- **Stability:** Agent's performance can vary significantly due to the stochastic nature of the environment and learning process.

- **Overfitting:** The agent may learn a policy that performs well on a specific instance but fails to generalize.

Outputs

Variation 1	**Variation 2**	**Variation 3**
Replay capacity: 10,000	Replay capacity: 1000	Replay capacity: 5,000
Batch size: 64	Batch size: 128	Batch size: 16
Epsilon: 0.9	Epsilon: 0.99	Epsilon: 0.5
Epsilon decay: 0.995	Epsilon decay: 0.999	Epsilon decay: 0.9
Gamma: 0.95	Gamma: 0.99	Gamma: 0.5
Learning rate: 0.001	Learning rate: 0.005	Learning rate: 0.1

Outputs