LECTURER: TAI LE QUY

INTRODUCTION TO REINFORCEMENT LEARNING

UNIT 5

DEEP REINFORCEMENT LEARNING

STUDY GOALS

- Understand the importance of neural networks in reinforcement learning
- Analyze and evaluate the deep Q-network algorithm and its components
- Apply knowledge of deep reinforcement learning to identify and evaluate scenarios for its use

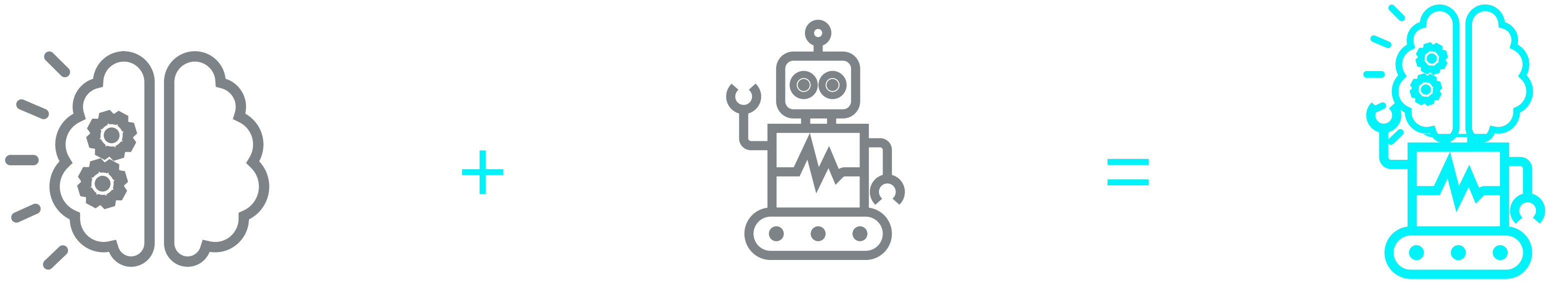
EXPLAIN SIMPLY

1. Explain the role of neural networks in improving the performance of reinforcement learning algorithms

2. Describe the components of deep Q-learning algorithm and how they contribute to its success?

3. Compare and contrast deep reinforcement learning with traditional reinforcement learning

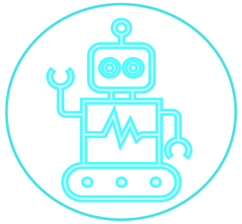
Synergy of neural networks and reinforcement learning



Neural Network + Reinforcement Learning = Deep Reinforcement Learning

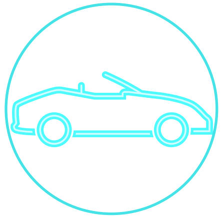
Games

- Atari games (pong, pacman)
- Go



Robotics

- Locomotion
- Grasping and navigation



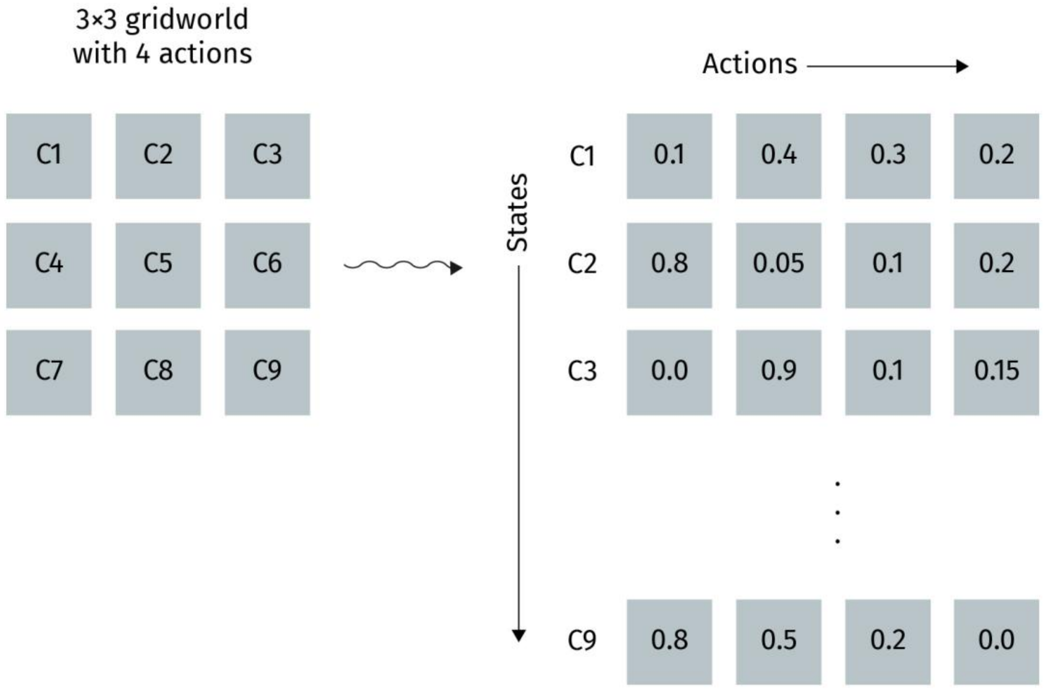
Autonomous Driving

- CARLA
- Navigation tasks

Deep Reinforcement Learning

Q-learning: simple and effective RL algorithm

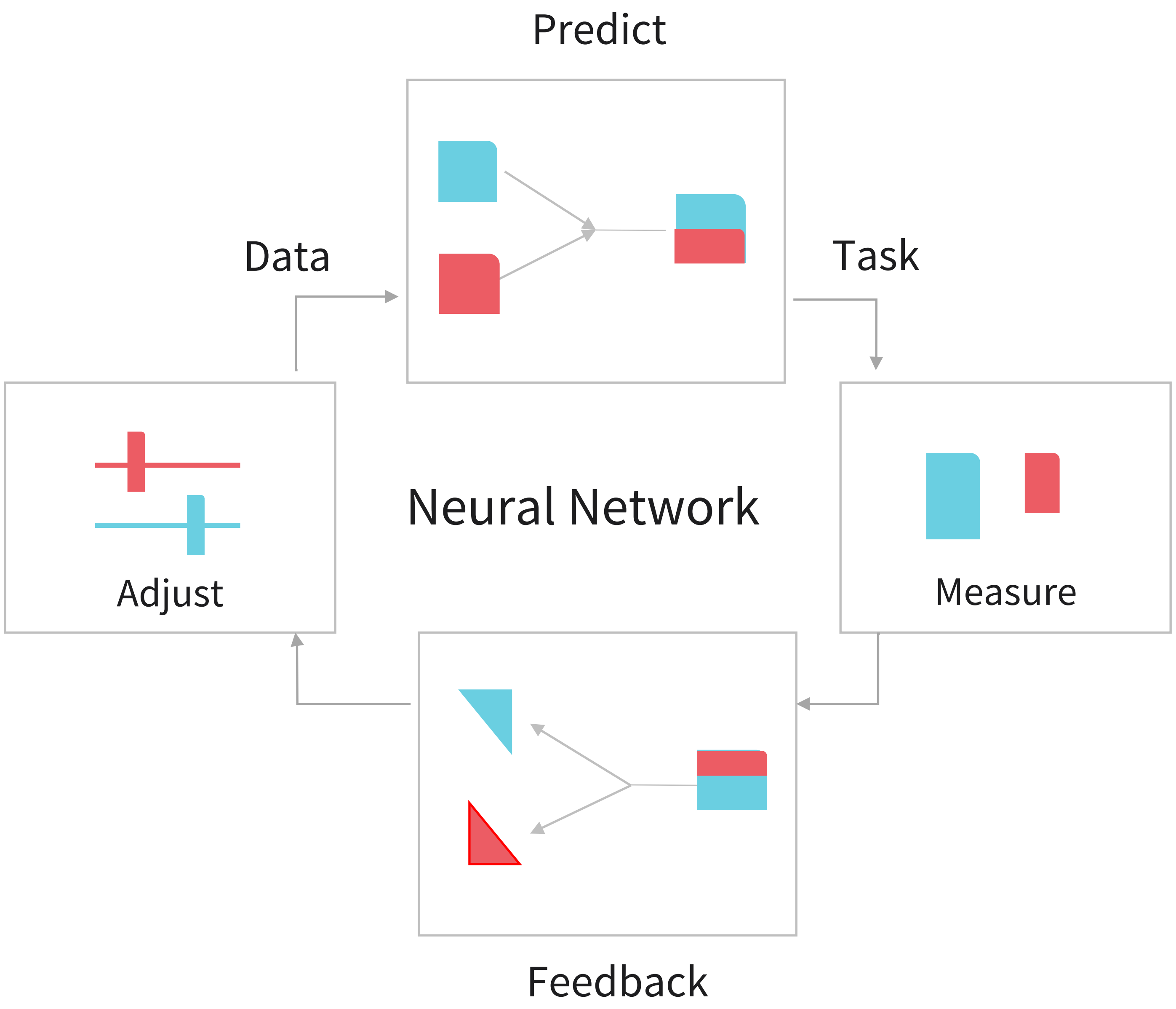
- Builds a lookup table
- Determines action to perform in state
- Guaranteed convergence (given enough steps)
- **Not scalable** for complex problems



Source of the image: Nair, 2023.

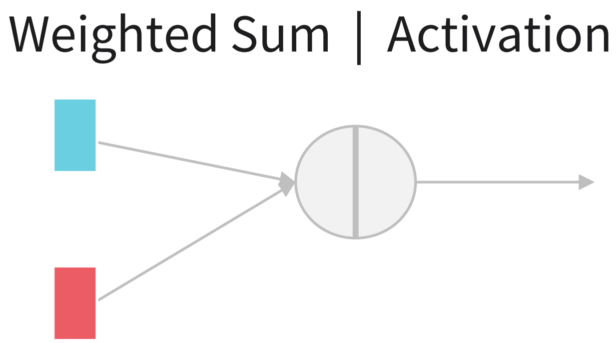
FUNDAMENTAL ELEMENTS OF NEURAL NETWORKS

- Map features to target
- Train and test
- Adjust and repeat



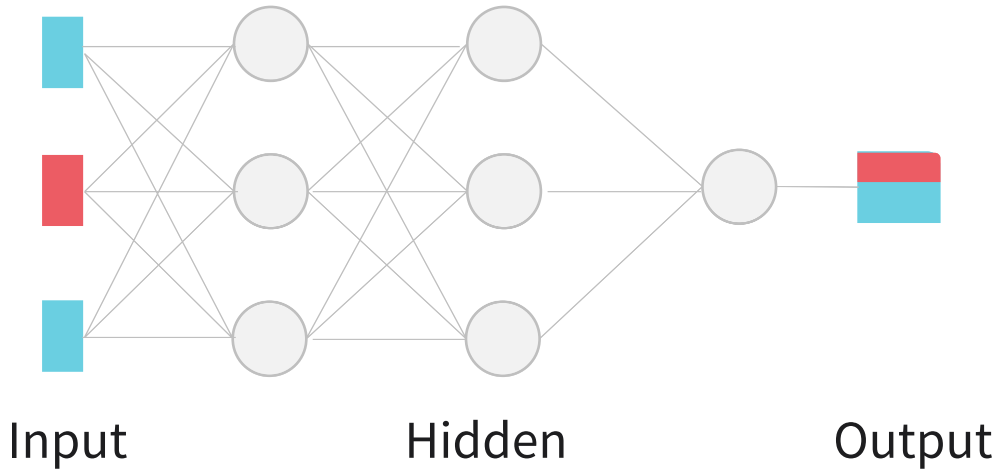
Source of the image: Plaku, 2023.

- Neuron: core computing unit
- Layers: receive input and transform it
- Model: combine layers into network



Artificial Neuron

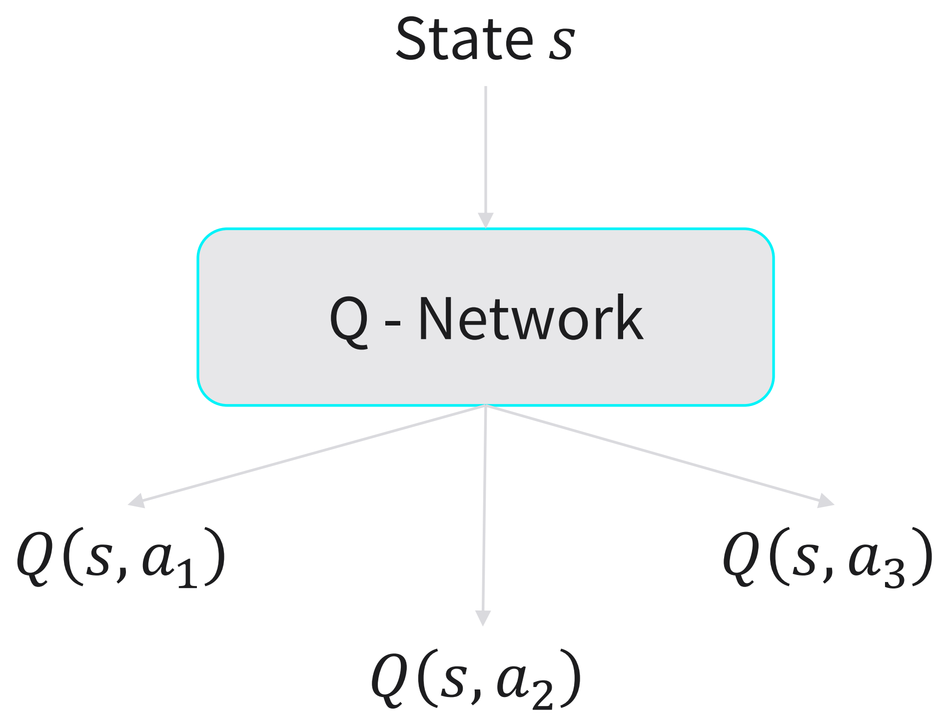
Neural Network



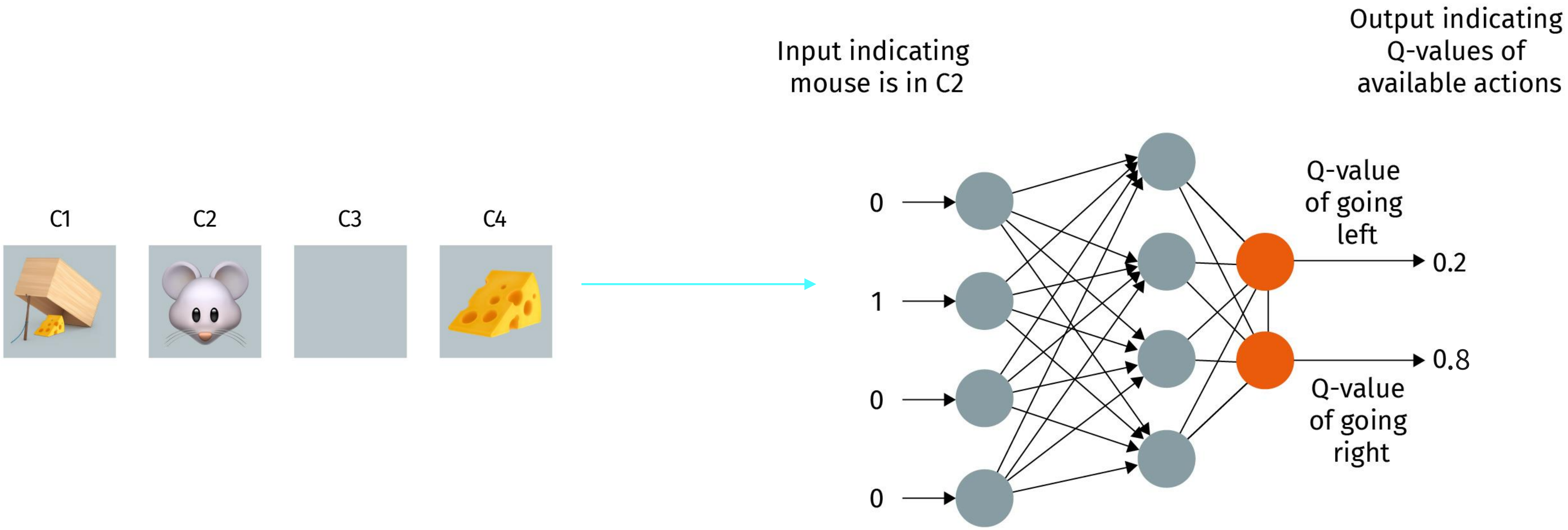
Source of the image: Plaku, 2023.

Efficient Q-value computation for complex environments

- Use neural networks instead of Q-table
- Goal: learn to compute Q-value for state-action pair $Q(s_t, a_t) \approx Q(s_t, a_t; \theta)$
- Tackle a wide range of problems and formats (image, video, text, etc.)



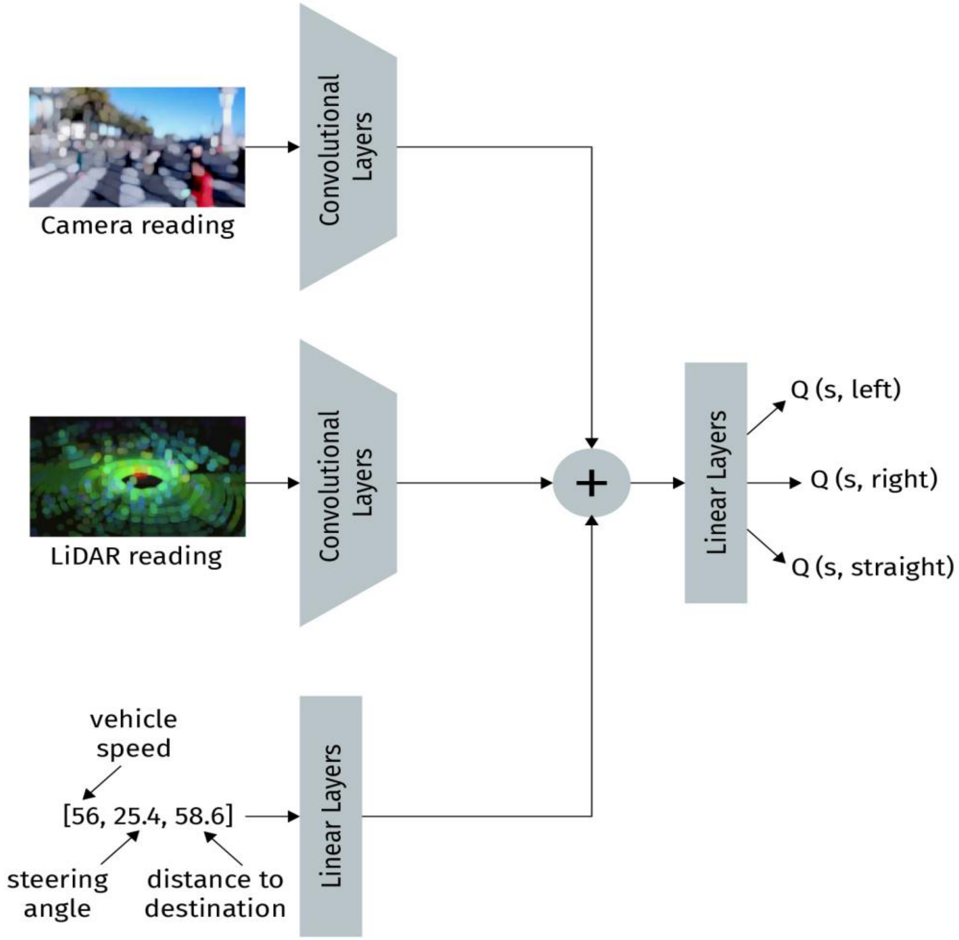
Goal: move the mouse to help it reach the destination



Source of the image: Nair, 2023.

Leverage sensor readings and neural networks to drive autonomously

- Combination of sensors
- Several neural networks
- Goal: learn steering commands to reach goal while avoiding obstacles



Source of the image: Nair, 2023.

Reframe reinforcement learning as supervised learning problem

- Approximate Q-function to update Q-values
- Learn best Q-value approximation
- Use temporal differences for target Q-values

$$Q(s_t, a_t; \theta) \approx r_t + \gamma \, . \, max_{a_{t+1}} Q(s_{t+1}, a_{t+1}; \theta)$$

θ: the parameters of the network

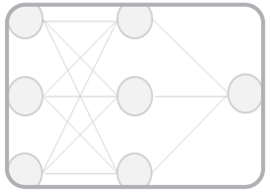
$$\text{Loss per data sample} = \Big(\Big(r_t + \gamma \, . \, \max_{a_{t+1}} Q\Big(s_{t+1}, a_{t+1}; \theta\Big)\Big) - Q\Big(s_t, a_t; \theta\Big)\Big)^2$$

- Use optimization algorithm to train neural network

OPTIMIZING DEEP REINFORCEMENT LEARNING

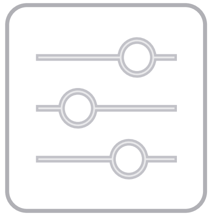
Replay buffer

- Store experience in replay buffer
- Sample **mini-batch** to update weights
- Improve efficiency and stabilize learning



Target network

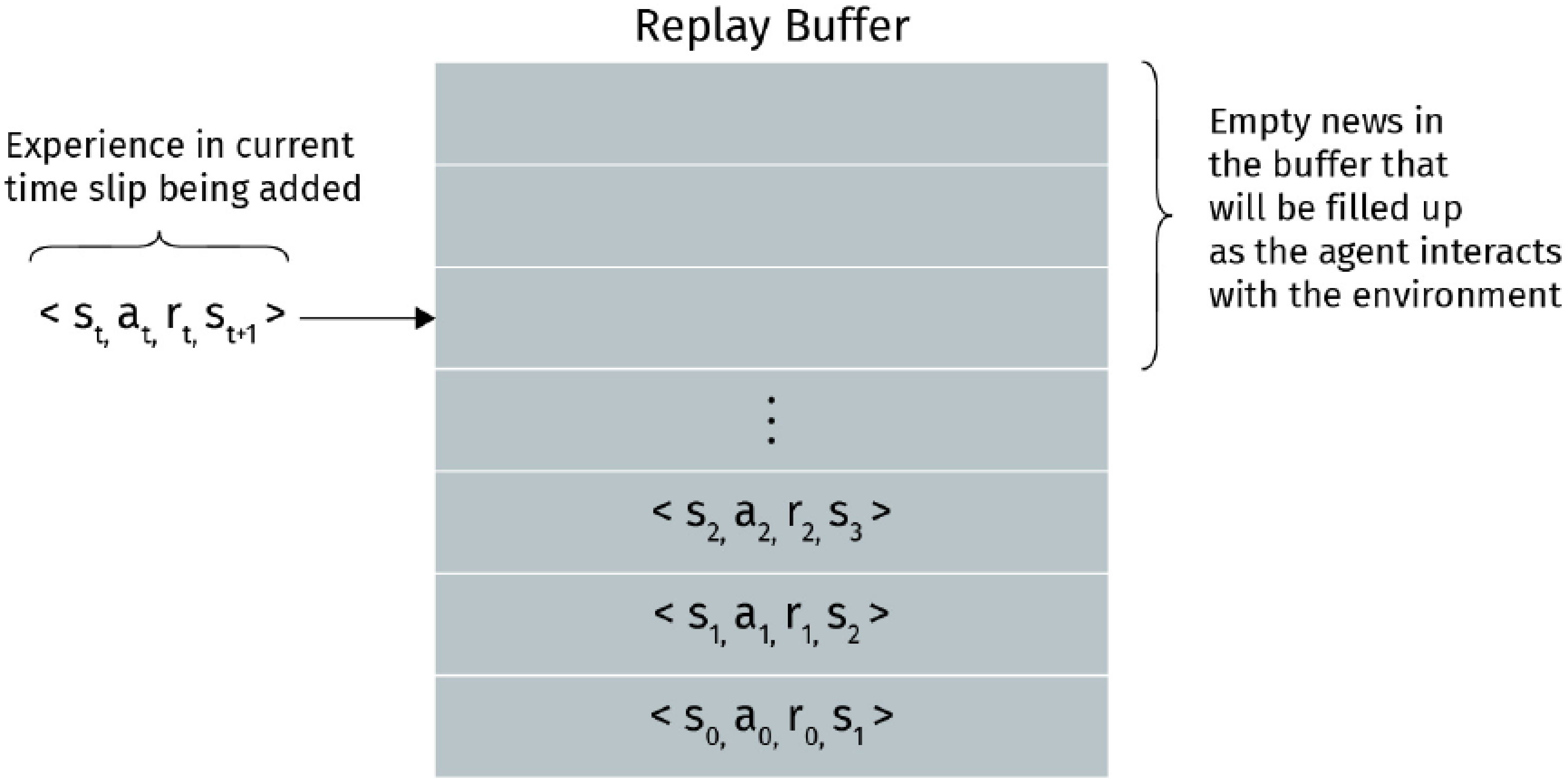
- Non-stationary targets in RL
- Target duplicates main network
- Target lags behind a few time steps



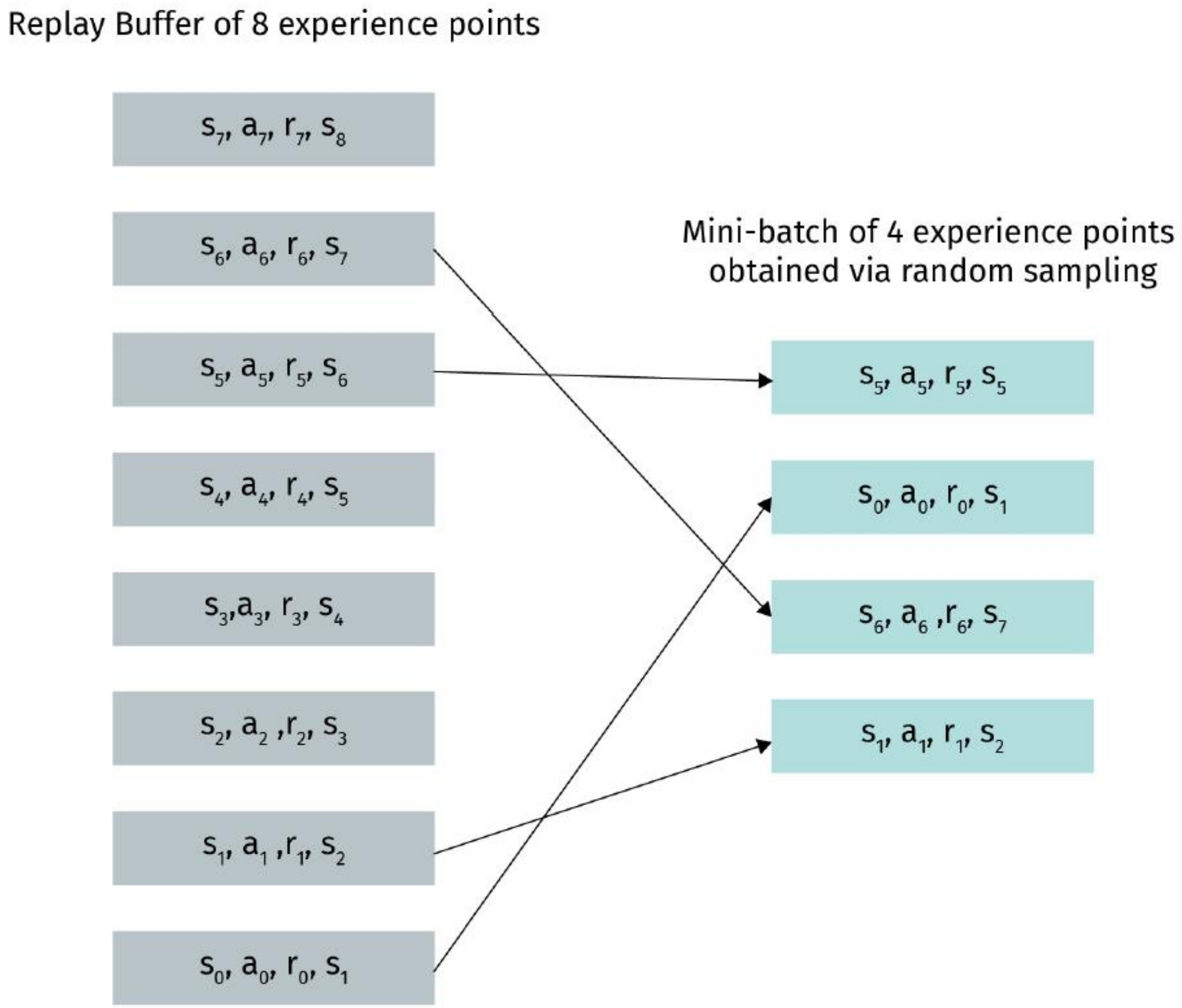
Network training

- Agents balance exploration vs exploitation
- Combine Bellman equations with greedy-epsilon strategy

REPLAY BUFFER



REPLAY BUFFER



REVIEW STUDY GOALS

- Understand the importance of neural networks in reinforcement learning
- Analyze and evaluate the deep Q-network algorithm and its components
- Apply knowledge of deep reinforcement learning to identify and evaluate scenarios for its use

SESSION 5

TRANSFER TASK

Case study

Your task is to develop an AI-agent that can play the Atari game Pong using deep Q-learning. Your goal is to develop an optimal policy that maximizes the game score

Task

What key steps will you consider when implementing Q-learning? Which neural network architecture would be most appropriate? How to optimize Q-learning? Consider factors such as experience replay and target network updates, among others

TRANSFER TASK
PRESENTATION OF THE RESULTS

Please present your results.
The results will be discussed in plenary.

1. The neural networks in DQN estimate
   a) State value (V-value)
   b) Action value (P-value)
   c) Advantage value (A-value)
   d) State-action value (Q-value)

2. The target network is introduced to resolve
   a) The non-IID nature of the data
   b) The non-stationarity of targets
   c) Sample inefficiency of RL algorithms
   d) Break correlations between experience tuples

LEARNING CONTROL QUESTIONS

3. The replay buffer is used to store
   a) Q-values
   b) Reward functions
   c) Experience tuples
   d) Policies

LIST OF SOURCES

Images

Plaku, 2023.
Nair, 2023.