# REINFORCEMENT LEARNING AGENTS FOR OPTIMAL PORTFOLIO MANAGEMENT

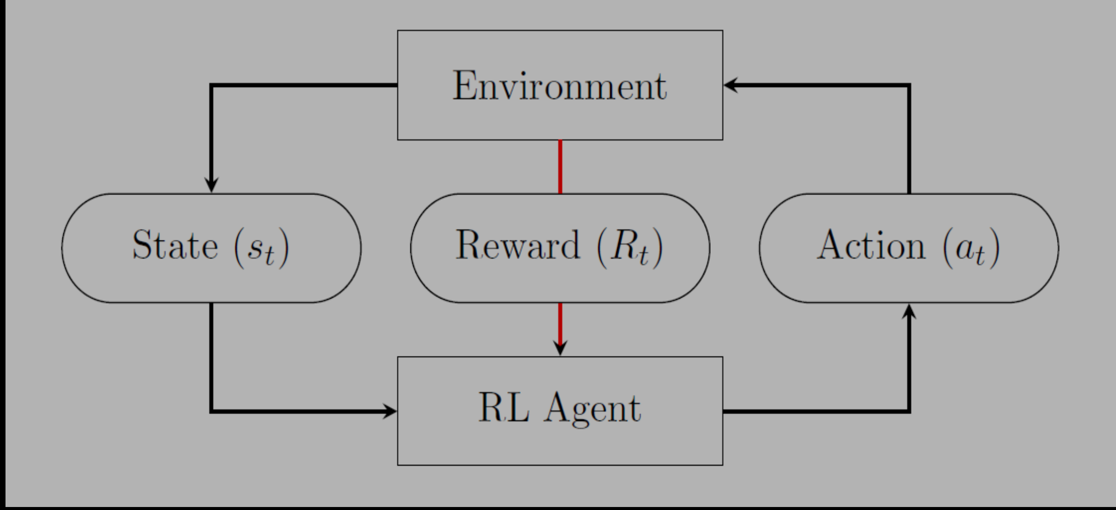
# Outline

- Portfolio Optimization
- Reinforcement Learning
- Problem Statement
- Scope
- Aims and Objectives
- Data and Tools

# Portfolio Optimization

- *Modern Portfolio Theory – Markowitz (1952)*
  - For a portfolio of multiple assets, maximize *expected* returns - based on a given level of market risk
  - An investment's risk and return characteristics should be evaluated by how the investment affects the overall portfolio's risk and return
  - Conversely, given a desired level of expected return, an investor can construct a portfolio with the lowest possible risk
  - Risk approximated from asset variance and correlation

# Reinforcement Learning




- (State) Value: *the expected cumulative reward from a state* $s_t$
- Policy: *what action to take in a given state.*
  - Defined as a mapping from state to actions
- Goal: Learn the policy that maximizes (cumulative) reward
- (Fully observable) Markov Decision Process
  - $P(r_{t+1}, s_{t+1} | s_t, a_t)$
  - Markov assumption: “The future is independent of the past, given the present”
  - $s_t$ = environment state

# Reinforcement Learning: Categories

- Categorizing RL agents:
  - *Value Based/Q Learning (DQN, C51, HER)*
  - *Policy Based (A2C, PPO)*
  - *Actor Critic (DDPG, TD3, SAC)*

  - *Model Free*
  - *Model Based (Dreamer, SimPLe)*

# Reinforcement Learning: DQN

“Playing Atari with Deep Reinforcement Learning” (Deepmind)

- Value based, off policy network
- First implementation to address instability arising from using nonlinear function approximators (eg. NNs) to learn the action-value function.
- 2 breakthroughs:
  - *Experience Replay using replay buffer of transitions,* $(s_t\ a_t\ r_{t+1}\ s_{t+1})$
  - *(Separate) Target Network*

# Reinforcement Learning: DDPG

“Continuous Control with Deep Reinforcement Learning” (Deepmind)

- Actor critic, off policy agent
- Extends the DQN approach to DPG
- Also uses replay buffer
- DQN works well for a high-dimensional discrete observation spaces, but cannot handle continuous high-dimensional action spaces

# Problem Statement

- To build RL agents that *dynamically* optimize the allocation of wealth towards assets in a portfolio in order to maximize returns for a given level of risk.

# Scope

- Overview of RL and its use in finance
- Introducing the portfolio optimization problem
- Literature Review of RL agents used in portfolio management
- Present the DQN and DDPG algorithms with their background, advantages, use cases, etc
- Implementation of the algorithms for portfolio management
- Discussion on results obtained

# Aims and Objectives

- Build a deep understanding of Reinforcement Learning, and its components, including the differences b/w value iteration and policy iteration, discrete and continuous action spaces, off-policy and on policy models, etc.
- Understand the use cases and applications of RL - especially in finance, and the challenges associated with the task of portfolio optimization.
- Through their implementation, gain experience and understanding on the technical details of the DQN and DDPG algorithms and the tools for building RL agents/environments.
- Demonstrate their usefulness at the task of portfolio management and compare performance.

# Data and Tools

https://gym.openai.com

https://stanford.edu/class/engr108/portfolio.html

# To-do

- Familiarization with the tools
- Basic implementations of DQN and DDPG architectures
- Literature Review

*FIN*