Machine Learning and Financial Applications

Lecture 10
Reinforcement Learning for Portfolio Optimization

Liu Peng

liupeng@smu.edu.sg

SMU Classification: Restricted

Video tutorial

- https://youtu.be/Yd3H_h-7C58

Introducing portfolio optimization

- The foundations of portfolio optimization can be traced back to the seminal work of Harry Markowitz, who introduced the concept of Modern Portfolio Theory (MPT) in his 1952 seminal paper Markowitz (1952)
- The essence of MPT lies in the quantification of the trade-off between risk and return, both at portfolio level.
- This leads to the formation of the efficient frontier, which represents a set of optimal portfolios offering the highest expected return for a given level of risk, or equivalently, the lowest risk for a target expected return.

Expected return and risk of the portfolio

$$E[R_p] = \boldsymbol{w}^T \boldsymbol{\mu}$$

$$\sigma_p^2 = \boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w},$$

$$\boldsymbol{w} = [w_1, w_2, \ldots, w_n]^T$$

Understanding asset return and covariance matrix

$$r_{t+1} = \frac{P_{t+1} - P_t}{P_t} = \frac{P_{t+1}}{P_t} - 1.$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix},$$

$$\mu_i = \frac{1}{T}\sum_{t=1}^{T} r_{i,t}. \qquad \mathrm{Cov}(r_i, r_j) = \sigma_{ij} = \frac{1}{T-1}\sum_{t=1}^{T}(r_{i,t} - \mu_i)(r_{j,t} - \mu_j),$$

Mean-variance optimization

$$
\begin{aligned}
\underset{\boldsymbol{w}}{\text{maximize}} \quad & \boldsymbol{w}^T \boldsymbol{\mu} \\
\text{subject to} \quad & \boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w} \leq \sigma_p^2, \\
& \boldsymbol{w}^T \mathbf{1} = 1,
\end{aligned}
$$

$$
\begin{aligned}
\underset{\boldsymbol{w}}{\text{minimize}} \quad & \boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w} \\
\text{subject to} \quad & \boldsymbol{w}^T \boldsymbol{\mu} = \mu_p, \\
& \boldsymbol{w}^T \mathbf{1} = 1,
\end{aligned}
$$

DRL for portfolio optimization

- Model-free learning
- Adaptive strategies
- Complex decision-making
- Scalability

Exploration versus exploitation

Markov decision process

- An MDP is a mathematical framework that provides a formal description of an environment for RL.

$$(S, A, P, R, \gamma)$$

$$s_t = \begin{bmatrix} P_t \\ I_t \\ \boldsymbol{w}_{t-1} \end{bmatrix}$$

$$a_t = \boldsymbol{w}_t = \begin{bmatrix} w_{1,t} \\ w_{2,t} \\ \vdots \\ w_{N,t} \end{bmatrix}, \quad \text{subject to} \quad \sum_{i=1}^{N} w_{i,t} = 1.$$

$$r(s_t, a_t) = \log\left(\sum_{i=1}^{N} w_{i,t} \frac{P_{i,t}}{P_{i,t-1}}\right).$$

$$p(s_{t+1}|s_t, a_t) = \mathbb{P}(S_{t+1} = s_{t+1}|S_t = s_t, A_t = a_t),$$

$$\pi^* = \arg\max_{\pi} \mathbb{E}\left[\sum_{t=0}^{T-1} \gamma^t r(s_t, \pi(s_t))|s_0\right],$$

The learning environment cont'd

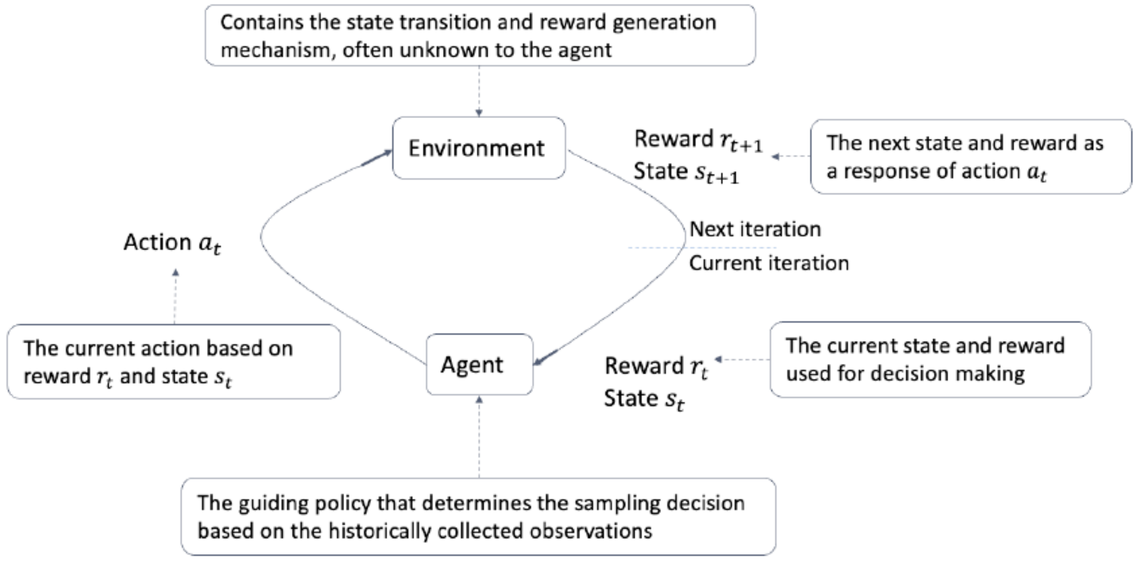

Figure 1.1: Iterative interaction between the agent and the environment.

Value functions

$$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \gamma^3 R_{t+4} + \cdots$$

$$G_t = R_{t+1} + \gamma G_{t+1}$$

$$V^\pi(s) = \mathbb{E}_\pi[G_t | S_t = s]$$

$$V^\pi(s) = \mathbb{E}_\pi[R_{t+1} | S_t = s] + \gamma \mathbb{E}_\pi[G_{t+1} | S_t = s]$$

$$V^\pi(s) = \sum_a \pi(a|s) \sum_{s',r} p(s', r | s, a)[r + \gamma V^\pi(s')]$$

$$Q^\pi(s, a) = \mathbb{E}_\pi[G_t | S_t = s, A_t = a]$$

$$Q^\pi(s, a) = \sum_{s',r} p(s', r | s, a)[r + \gamma \sum_{a'} \pi(a'|s') Q^\pi(s', a')]$$