

DMUU Assignment 2 – Reinforcement Learning in Finance

Academic integrity – read this before you start

This assignment emphasizes independent analysis and interpretation. You may engage in general discussion with classmates about the readings, but:

- All answers, drawings, and written analyses must be your own work.
- The use of AI tools is permitted for this assignment. However, you must demonstrate your own reasoning, critical thinking, and originality throughout.

Time expected	About 3-4 hours: 1.5 hours reading, 1.5 hours for the questions
Full pass	Submitted on Moodle by Monday 19 October 2026, 12:00
Half pass	Submitted between Monday 12:00 and Tuesday 20 October, 10:00
No credit	After Tuesday 20 October, 10:00
File	PDF, named LastName_FirstName_Assignment2.pdf

How to pass

The assignment is pass/fail. You pass if you:

1. Complete Tasks 1, 2 and 3 in Part 2: Complete the backup diagram. Annotate states and transitions, answer the questions
2. Part 1 is a reading exercise. We encourage you to discuss the article and the guiding questions. You don't have to submit answers to the guiding questions.

What we check, and what you check

Pass/fail works both ways: you need NOT polish every answer, and we will not correct every answer. Graders check only whether the pass criteria are met and stop once they are, so you will not receive comments on individual answers.

After the final cutoff (Tuesday 10:00) we publish an answer guide on Moodle (after class). Compare your answers with it yourself. If something is unclear, ask in class or on the Moodle forum.

Part 1. Guided Reading.

Reinforcement Learning, and AI in general, take a more and more important role in Finance. The following link plus the Medium-article give you an idea of the different real-world use-cases.

Read them. We encourage you to discuss the role of AI in finance with your colleagues. For guidance, we provide a few questions below.

<https://www.finextra.com/newsarticle/36761/rbc-launches-ai-guided-electronic-trading-platform>

"RBC launches AI-guided electronic trading platform": ... *"Built with the Canadian bank's Borealis AI research centre, the platform, called Aiden, looks to apply deep reinforcement learning into a constantly changing environment like equities trading."*

<https://medium.com/@liangnguyen612/reinforcement-learning-in-finance-a-practitioners-roadmap-2b84d84686a6>

"Reinforcement Learning in Finance: A Practitioner's Roadmap",

Quang Khải Nguyễn Hưng, Medium
(PDF on Moodle)

Guiding Questions:

1. Why are nearly all RL trading systems trained in simulators before they trade real money? (Section 1)
2. Why does the author call finance "almost perfectly structured as an RL problem"? (Section 1)
3. Why do many impressive results of RL portfolio strategies disappear when tested on later periods? (Section 6e)

Part 2. The scenario: Invest or keep cash

In this assignment we study a simplified investment decision and model it in a way a Reinforcement Learning agent could learn the optimal policy (strategy) from experience:

An agent has 100 CHF for two years. It always holds all its money either in cash or in a fund.

Rules

- At the beginning of year 1, the agent chooses: buy the fund, or keep cash.
- At the beginning of year 2, it chooses again. In the fund, it can hold or sell. In cash, it can buy the fund or keep cash.
- Each year the fund gains 15 CHF (probability 60%) or loses 10 CHF (probability 40%). The amounts (+15 or -10 CHF) are the same in both years (not percentages).
- Cash earns 0.5 CHF interest per year.
- Each trade (buy or sell) costs 1 CHF.
- At the end of year 2, everything is sold. This is the terminal state.

Reward. After each action, the agent receives a reward R :

- R = interest received + gain or loss when the fund is sold - trading costs
- A gain or loss counts only when the fund is sold. Buying therefore gives $R = -1$.
- Example from the diagram: buy, fund up, hold, fund up. The rewards are $R = -1$ (in the first step), then $R = 130 - 100 - 1 = +29$ (in the second step). The final wealth is 128 CHF.

States show what the agent holds and its total wealth. One could model multiple terminal states, but here we decided to have only one. The wealth at the end of each possible path is written along the arrow next to **"Final"**.

Check: final wealth = 100 + sum of R along the route.

We do not consider discounting (set $\gamma=1$).

Task 1. Complete the backup diagram on page 5. Part of it is given, add:

- the missing arrows from the black dots, each with R and its probability (or "certain"),
- the total wealth in "Fund down" and "Cash", with the calculation, as in "Fund up",
- the two actions of "Fund down" and of "Cash", each with R , probability and final wealth.

Task 2. Q-values and the optimal policy. Given the diagram we can derive the Q-values from it (an agent does not know it! see next task):

- Consider the situation that the agent kept the cash in year 1, and then decides to buy the fund in year 2. That is, compute the value $Q(\text{state}=\text{"cash"}, \text{action}=\text{"buy"})$. Note that this state/action pair has a stochastic output: the fund can go up with a probability of 60%, or down with a probability of 40%. The Q-value here is the weighted sum of the two possible outcomes ($0.6 \times R_{\text{up}} + 0.4 \times R_{\text{down}}$).

- b) Calculate all Q-values in the diagram
Note: for a non-terminal state, sum not only the immediate rewards, but all expected future rewards. Assume the agent takes the better action in year 2.
- c) What is the best action in the first year?

Task 3. Learning from experience. An RL agent does not know the diagram. It has to learn the Q-values from experience. Here, we do not do a calculation of the SARSA updates, but study RL conceptually:

- a) How many Q-values does the agent have to learn in our scenario?
- b) Assume the agent must try each action **a** in each state **s** about 10 times before the estimates of $Q(\mathbf{s}, \mathbf{a})$ are reliable. One episode is two years of investing. Roughly how many years of investing would the agent need? Are 20 years of historic market data enough?
- c) Real RL systems in finance get their experience from historical market data or from simulated markets. Give one advantage and one risk of each.
- d) In algorithmic trading, agents place orders within fractions of a second. Is learning from experience easier or harder for such an agent than in our scenario? Explain.

Backup diagram (incomplete):

