Technical University of Crete
School of Electrical and Computer Engineering
Reinforcement Learning and Dynamic Optimization

Project: Stock Portfolio Optimization
Phase 1 - Assignment 3

Nikolaos Angelidis (ID: 2019030190)
Georgios Gialouris (ID: 2019030063)

Investment Rules (Environment):

- Each day (round) we invert exactly 1 euro. At the end of the day we receive as profit the percentage increase of the respective stock. Our goal is to maximize the cumulative profit in the long run.
- If we choose to invest in the same stock, then no transaction fee must be paid. However, if we decide to switch stocks, then a flat transaction fee of c euros must also be paid ($c \in [0,1]$).
- Unless otherwise stated, we keep playing this game for an infinite horizon T and inflation is captured in the discount factor γ (the higher it is the smaller the inflation).

Stock Price Evolution (Environment): There are N stocks available, and the price of each stock i evolves as a 2-state Markov Chain, between a "H(igh)" and a "L(ow)" price as follows:

- When at state H, the expected gain of the stock is r_i^H. With probability p_{HH}^i the stock will remain in state H, and with probability p_{HL}^i the stock will go to state L.
- When at state L, the expected gain of the stock is r_i^L. With probability p_{LL}^i the stock will remain in state L, and with probability p_{LH}^i the stock will go to state H.

Control Variable and Objective (Agent):

- In each round, our algorithm must choose which stock to invest in.
- Our objective is to maximize the expected (discounted) long-term profits, starting from any state.

This is a Markov Decision Process (MDP) problem, which means that our agent knows all the environment parameters (the current state of every stock, the transition probabilities and expected rewards at each state, etc.).

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

State Space: In order to be able to decide what is the best action to take every time, our state should include the following:
i) the Markov state (H or L) for each stock, and
ii) Which stock we are currently holding.
The latter is needed as it affects whether we will incur a transaction fee or not. The former is needed in order to know the correct probability of each stock giving a H(igh) or L(ow) return in the next round.

So, for $N=2$, the possible states are

$$S=\{\{1,H,H\},\{1,H,L\},\{1,L,H\},\{1,L,L\},\{2,H,H\},\{2,H,L\},\{2,L,H\},\{2,L,L\}\},$$

a total of $|S|=8$ states.

For N stocus, the possible states are

$$S=\{\{1,H_1,\dots,H_N\},\dots,\{1,L_1,\dots,L_N\},\dots,\{N,H_1,\dots,H_N\},\dots,\{N,L_1,\dots,L_N\}\},$$

a total of $|S|=N\cdot 2^N$ states.

Action Space: At any state, we have N available actions. Each action represents a stocu i in which we can choose to invest in the next round.

So, for $N=2$, the available actions are

$$A=\{1,2\}$$

For N stocus, the available actions are

$$A=\{1,\dots,N\}$$

Transition Probabilities: The transition probabilities only depend on the 2-state Markov chains. The actual action will affect only the first component of the state (i.e., which stocu we will have in the next round).

For $N=2$, the transition probabilities $P(s,a,s')$ from state $\{1,H,H\}$ are:

$$P(\{1,H,H\},1,\{1,H,H\})=\rho^1_{HH}\cdot\rho^2_{HH}=P(\{1,H,H\},2,\{2,H,H\})$$

$$P(\{1,H,H\},1,\{1,H,L\})=\rho^1_{HH}\cdot\rho^2_{HL}=P(\{1,H,H\},2,\{2,H,L\})$$

$$P(\{1,H,H\},1,\{1,L,H\})=\rho^1_{HL}\cdot\rho^2_{HH}=P(\{1,H,H\},2,\{2,L,H\})$$

$$P(\{1,H,H\},1,\{1,L,L\})=\rho^1_{HL}\cdot\rho^2_{HL}=P(\{1,H,H\},2,\{2,L,L\})$$

The transition probabilities $P(s,a,s')$ from the remaining 7 states are obtained in a similar way.

For N stocus, the transition probabilities $P(s,a,s')$ from state $\{1,H_1,\dots,H_N\}$ are:

$$P(\{1,H_1,\dots,H_N\},1,\{1,H_1,\dots,H_N\})=\rho^1_{HH}\cdot\dots\cdot\rho^N_{HH}$$

The transition probability remains exactly the same regardless of the action.

$$P(\{1,H_1,\dots,H_N\},1,\{1,L_1,\dots,L_N\})=\rho^1_{HL}\cdot\dots\cdot\rho^N_{HL}$$

...

The transition probabilities $P(s,a,s')$ from the remaining $N\cdot 2^N-1$ states are obtained in a similar way.

Rewards: The rewards depend only on the stock we will have, and its state (H or L). **Additionally**, we must subtract the transition fee, if the action we took was to switch stocks.

For N stocks, the rewards $R(s,a,s')$ are:

If the action is to invest in the same stock, then:

if the previous state of the stock i was H, then:

$R_t = r_i^H$ with probability p_{HH}^i and

$R_t = r_i^L$ with probability p_{HL}^i

if the previous state of the stock i was L, then:

$R_t = r_i^H$ with probability p_{LH}^i and

$R_t = r_i^L$ with probability p_{LL}^i

If the action is to switch stocks, then:

if the previous state of the new stock i was H, then:

$R_t = r_i^H - c$ with probability p_{HH}^i and

$R_t = r_i^L - c$ with probability p_{HL}^i

if the previous state of the new stock i was L, then:

$R_t = r_i^H - c$ with probability p_{LH}^i and

$R_t = r_i^L - c$ with probability p_{LL}^i

Terminal States: There are no terminal states in this problem.