QF624-2025-W4

Number of participants: 40

1. Which of the following is typically NOT a direct state feature in a Reinforcement Learning model for Portfolio Optimization?

12 correct answers
out of 25 respondents

	Current portfolio weights	40%	10 votes
	Moving averages	12%	3 votes
	Market sentiment	0%	0 votes
✓	Text data from news articles	48%	12 votes

2. What is the role of the action space in a Portfolio Optimization problem modeled using Reinforcement Learning?

11 correct answers
out of 21 respondents

	It represents the set of all possible market states.	5%	1 vote
	It represents the set of all possible rewards.	19%	4 votes
✓	It represents the set of all possible portfolios.	52%	11 votes
	It represents the set of all possible trading strategies.	24%	5 votes

3. What is the main challenge in using Reinforcement Learning for Portfolio Optimization?

10 correct answers
out of 23 respondents

	Overfitting	48%	11 votes
	High computational cost	9%	2 votes
✓	Non-stationarity of financial markets	43%	10 votes
	Lack of data	0%	0 votes

4. Which reward design would most directly encourage the agent to balance return and risk?

16 correct answers
out of 18 respondents

	Maximizing cumulative raw returns	6%	1 vote
	Minimizing turnover in the portfolio	0%	0 votes
✓	Maximizing a risk-adjusted return metric (e.g. Sharpe ratio)	89%	16 votes
	Penalizing any change in weight allocations	6%	1 vote

5. Which of the following is NOT a component of the Markov Decision Process (MDP)?

18 correct answers
out of 19 respondents

	States	0%	0 votes
	Actions	0%	0 votes
	Reward	5%	1 vote
✓	Clusters	95%	18 votes

6. What does the term 'exploration vs exploitation' refer to in Reinforcement Learning?

19 correct answers
out of 22 respondents

	The difference between model-based and model-free methods	14%	3 votes
✓	The trade-off between learning new strategies and using known strategies	86%	19 votes
	The choice between using value iteration and policy iteration	0%	0 votes
	The distinction between on-policy and off-policy methods	0%	0 votes

7. The optimal policy π^* is defined by

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

12 correct answers out of 14 respondents

	Immediate reward only	0%	0 votes
	Total undiscounted reward over horizon T	14%	2 votes
✓	Expected discounted sum of rewards	86%	12 votes
	Variance of returns	0%	0 votes

8. What role does the discount factor $\gamma \in [0, 1]$ play in the objective?

24 correct answers
out of 24 respondents

	It scales transaction costs in the reward	0%	0 votes
	It penalizes changes in portfolio weights	0%	0 votes
✓	It controls the trade-off between immediate and future rewards	100%	24 votes
	It defines the action-space constraints	0%	0 votes

9. **When the action is a portfolio weight vector $w_t \in \Delta^{N-1}$ (the probability simplex), one approach is to parameterize an unconstrained vector $y \in \mathbb{R}^N$ and set $w_i = \frac{\exp(y_i)}{\sum_j \exp(y_j)}$. This softmax mapping ensures $\sum_i w_i = 1$ and $w_i > 0$. What is a potential drawback of this parameterization in high-dimensional portfolios?**

15 correct answers
out of 20 respondents

	Softmax outputs can be negative	5%	1 vote
✓	Numerical instability when some $y_i \gg y_j$	75%	15 votes
	Loss of differentiability	15%	3 votes
	Violation of the simplex constraint	5%	1 vote

10. Which of the following statements is \textbf{true} regarding the relationship between $v^*(s)$ and $q^*(s,a)$?

14 correct answers
out of 18 respondents

	$v^*(s)$ is always equal to $q^*(s,a)$ for any action a	17%	3 votes
	$v^*(s) = \min_a q^*(s,a)$	0%	0 votes
✓	$v^*(s) = \max_a q^*(s,a)$	78%	14 votes
	$v^*(s)$ is not related to $q^*(s,a)$ under optimality	6%	1 vote

11. To penalize high turnover, one might define $r_t = \log\left(\sum_i w_{i,t}\frac{P_{i,t}}{P_{i,t-1}}\right) - \lambda\sum_i |w_{i,t} - w_{i,t-1}|$. The effect of increasing $\lambda > 0$ is to:

15 correct answers
out of 21 respondents

	Encourage more frequent rebalancing	10%	2 votes
✓	Discourage large changes in portfolio weights	71%	15 votes
	Increase sensitivity to market volatility	10%	2 votes
	Remove risk-adjustment from the reward	10%	2 votes

12. Financial markets shift over time. Which technique can help an RL agent adapt its portfolio policy online?

11 correct answers
out of 23 respondents

	Fixed-size replay buffer with uniform sampling	4%	1 vote
	Periodically resetting the policy to random initialization	39%	9 votes
✓	Prioritized replay that over-samples recent transitions	48%	11 votes
	Using only on-policy Monte Carlo updates	9%	2 votes