# Lecture 01: Bandits and MDPs

#cs #rl

108 plays · 185 players

A public kahoot

## Questions (10)

1 - Quiz

**Exploration vs. Exploitation: What is correct?** 60 sec

- Exploitation: Take a random action. ✗
- Exploitation: Make best decision given current information. ✓
- Exploration: Try new actions. ✓
- Exploration: Apply best known action. ✗

2 - Quiz

**UCB can be categorized as ...** 60 sec

- Naive Exploration ✗
- Optimism in the Face of Uncertainty ✓

3 - True or false

**The more often an action has been chosen, the more likely it is chosen again by UCB.** 20 sec

- True ✗
- False ✓

4 - Quiz

**Bandit strategies for solving the exploration-exploitation issue for RL are suboptimal,...** 60 sec

- because they ignore the sequence of actions to be made. ✓
- because they are not an one-step decision-making approach. ✗
- because they are a multi-step decision-making approach. ✗
- because they consider the sequence of actions to be made. ✗

5 - Quiz

**In contextual bandits, ...** 60 sec

- we also consider the current state (i.e. context). ✓
- the best action depends on the given context. ✓
- the best action does not depend on the given state. ✗

6 - Quiz

**A Markov Decision Process is defined by a set of states and ...** 20 sec

- a value function. ✗
- transition probabilities between states. ✓
- a set of actions and a reward function. ✓
- a policy. ✗

7 - Quiz

**A discount factor smaller 1 is required in a Markov Decision Process for** 20 sec

- focusing on future rewards. ✗
- taking uncertainty about the future into account. ✓
- avoiding infinite rewards. ✓
- ensuring infinite rewards in the limit. ✗

8 - Quiz

**The state value function v(s) of an MDP is** 20 sec

- ✗ the expected return starting from state s and a specific action a.
- ✓ the expected return starting from s given a policy pi.

9 - Quiz

**Which one is correct?** 60 sec

- ✗ (1)
- ✗ (2)
- ✗ (3)
- ✓ (4)

10 - Quiz

**To "solve" an MDP, we have to determine ...** 20 sec

- ✗ the state-value function for an arbitrary policy.
- ✗ the action-value function for an arbitrary policy.
- ✓ the state-value function for an optimal policy.
- ✗ the unique optimal policy.