

**CS3018 Reinforcement Learning**fast-logo

**Quiz 2**

Date: 17th February, 2025

Give short answers to the following questions:

1. What is meant by optimistic initial values in multi-armed bandits problem and how is it useful for the agent?
2. In the Upper Confidence Bound algorithm, how an action at time t is chosen? Write down the equation for action A at time t i.e., At. Explain the equation.
3. How is the Markov Decision Process different from the Multi-armed bandits problem and what does the term Markov mean in MDP?
4. Consider the equation for updating the estimate of an action as:

Give the value of considering the following figure in which the lighter line is the target and the darker line is our estimate plotted over time.

