

**CS3018 Reinforcement Learning**fast-logo

**Quiz 5**

Date: 21st April, 2025

## **Question 1:**

## You are given the following experience collected by an agent in a deterministic environment: Time | State S | Action A | Reward R | Next State S' | Next Action A' --------------------------------------------------------------- 0 | A | right | 0 | B | down 1 | B | down | 0 | C | right 2 | C | right | 1 | Terminal | Initial Q-values are all 0. Using learning rate alpha = 0.5, compute: 1. The updated Q-values using SARSA 2. The updated Q-values using Q-learning

## **Question 2:**

An agent estimates the state-value function V(s) using TD(0). The agent starts in state S1, receives reward R=2, and transitions to state S2, where the estimated value is V(S2) = 5. If V(S1) = 3, and learning rate alpha = 0.1:  
  
Calculate the updated value of V(S1) using the TD(0) formula.

## **Question 3:**

Suppose after real interaction, your agent stores the following transition in its model:  
(S, A) -> (S', R = 1)  
The current Q-values are:  
 Q(S, A) = 0.5, Q(S', any) = 1.0  
  
Using a learning rate alpha = 0.5 and gamma = 0.9, perform one planning update on Q(S, A) using Dyna-Q's simulated experience:  
Show the updated value of Q(S, A)