

# **ECS/DSE-427/627: Multi-Agent Reinforcement Learning**

## **Midsem**

### **Question 1:**

**(30 marks)**

You are tasked with modelling and solving a modified version of the Travelling Salesman Problem (TSP) as a reinforcement learning (RL) problem. In this version of TSP, each target has an associated profit value that decays linearly with time such that,

$$p^i = p^i - \text{dist\_so\_far}()$$

where  $p^i$  is the profit of the target and  $\text{dist\_so\_far}()$  is the distance travelled.

There are a total of 10 targets. The locations of the targets are fixed, but the profit values are shuffled across the targets after every  $k$  episodes. Your task is to visit each target in a sequence to maximize the total collected profit.

The modified TSP code has been provided in the [marl-ecs-course](#) repository under `Midsem`.

Your task is to solve this problem with any TD-based reinforcement learning algorithm of your choice. (SARSA, Q-learning, DQN, SAC, etc..)

### **Key Deliverables:**

- Plot the episode vs cumulative reward plot to show the training convergence.
- The deadline to submit the assignment is until October 1st, 2:59 PM.
- You will be required to run the codes and test out your algorithm on a unit test configuration that will be provided soon.