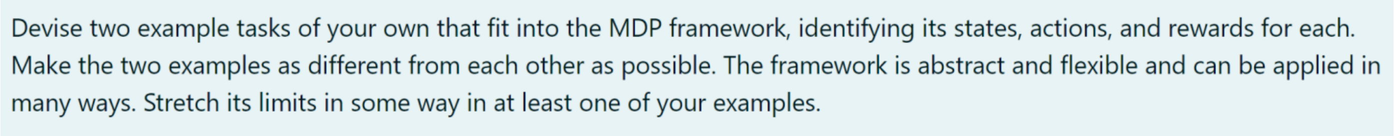
EE932 Assignment-2 Solution

---

eMasters in Communication Systems, IITK
EE932: Introduction to Reinforcement Learning
Instructor: Prof. Subrahmanya Swamy Peruru
Student Name: Venkateswar Reddy Melachervu
Roll No: 23156022

Question 7:



Devise two example tasks of your own that fit into the MDP framework, identifying its states, actions, and rewards for each. Make the two examples as different from each other as possible. The framework is abstract and flexible and can be applied in many ways. Stretch its limits in some way in at least one of your examples.

Solution:

Task 1: Drone Delivery System (Deterministic Policy)

Task Description

Imagine a drone delivery system where drones are tasked with delivering packages from a warehouse to various locations in a city. The objective is to minimize delivery time and energy consumption.

MDP Framework

States (S):

- Drone's current position, battery level, the location of package source, and the location of undelivered package(s).

Actions (A):

- Drone's movements between locations and the decision to pick up or drop off packages.

Rewards (R):

- Successfully delivering a package: +1 reward
- Taking the shortest route to the destination: Small positive reward.
- Taking the shortest route to the next source/pick-up location: Small positive reward.
- Consuming battery: Small negative reward proportional to battery usage.

Deterministic Policy:

- Predetermined path calculated based on the shortest distances between source and delivery destinations, with deterministic actions at each state.

Task 2: Spectrum Allocation in 6G Networks (Stochastic Policy)

Task Description

Consider a scenario where a central network controller is responsible for allocating spectrum resources to multiple users in a 6G wireless network. The objective is to maximize network throughput while ensuring fairness among users.

MDP Framework

States (S):

- C current spectrum allocation and network conditions, including signal interference and user demand.

Actions (A):

- Actions correspond to allocating different portions of the spectrum to users.

Rewards (R):

- High throughput achieved: Positive reward.
- Excessive interference: Negative reward.
- Fairness among users: Additional reward for maintaining fairness.

Stochastic Policy:

- The network controller employs a stochastic policy that probabilistically selects spectrum allocation actions based on the current state and historical data.
- The policy aims to balance exploration (trying out different allocation strategies) and exploitation (using known effective strategies).
- This stochastic policy allows the network controller to adapt to changing network conditions and user demands, considering uncertainty and variability in the wireless environment, which is essential for optimizing performance in future-generation wireless communication systems like 6G.

-- End of the Document --