Reinforcement Learning Workshop
Day 1 – Student Activities

Topics

- Markov Decision Processes (MDPs)
- Policy evaluation

---

1 Activity 1: Modeling an MDP?

Consider the decision-making problem of managing an inventory system for a single type of item. Each day a certain number of orders is received that must be shipped by the next day. The manager of the facility can, each day, order a number of items. The space available for inventory is limited, and if there are not enough items to fulfill outstanding orders a penalty is incurred.

The goal of the manager is to minimize the cost of operating the system in the long term. Discuss the following questions.

(a) What kind of costs would you expect to have in managing such system?

(b) What important parameters would have to be specified to model this problem?

(c) Is this an episodic task or a continuing task?

Taking into account the previous discussion, create an MDP model for this problem, specifying:

(a) The state space $\mathcal{S}$

(b) The action space $\mathcal{A}(s)$

(c) The transition probability matrices P^a for each action a.

(d) The reward matrices R^a for each action a.

(e) Is this a continuing task or an episodic task?

—

2 Activity 2: Episodic vs Continuing Tasks

Classify each task below as *episodic* or *continuing*.

- Playing a game of chess
- Controlling a thermostat
- Robot navigation with a goal state
- Stock portfolio management

For each task:

1. Does it have terminal states?
2. What is an appropriate discount factor γ?

—

3 Activity 3: Computing Returns by Hand

Consider the following trajectory:

$$(s_0, a_0, r_1 = 1),\ (s_1, a_1, r_2 = 2),\ (s_2, a_2, r_3 = -3),\ (s_3 \text{ terminal})$$

1. Compute the total return with $\gamma = 1$
2. Compute the total return with $\gamma = 0.9$

Question: How does the choice of γ change the importance of future rewards?

—

4 Activity 4: Evaluating a Policy

A MDP has state space $\mathscr{S} = 1, 2$ and action space , , . All actions are available in all states. The transition probability and reward matrices for each state are:

$$P^a = \begin{bmatrix} 0.2 & 0.8 \\ 0.7 & 0.3 \end{bmatrix} \quad R^a = \begin{bmatrix} 10 & 7 \\ 12 & 15 \end{bmatrix}$$

$$P^b = \begin{bmatrix} 0.4 & 0.6 \\ 0.1 & 0.9 \end{bmatrix} \quad R^b = \begin{bmatrix} 5 & 11 \\ 14 & 7 \end{bmatrix}$$

$$P^c = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix} \quad R^c = \begin{bmatrix} 14 & 3 \\ 2 & 12 \end{bmatrix}$$

Consider the deterministic policy $\pi(1) = c$, $\pi(2) = b$

(a) Write the Bellman equations for the state value function $V^\pi(s)$ for this policy.

(b) Write the equations in matrix form $AV^\pi = b$

(c) Solve linear system to find V^π (either by hand or using a computer).

5 Optional Coding Activity

Go to the GitHub repository for the course (`https://github.com/lfmartins/rl_cimpa_2026` and click on the "Day 1 Activities in Colab" link.

—