ECEN 743: Reinforcement Learning
Assignment 1

Overview

1. You will complete this assignment in a Jupyter notebook. You can download it from our (ECEN743 GitHub repo).
2. You have to submit an HTML report and your code to Canvas.
3. Put all your files (HTML report and code) into a **single compressed folder** named **LastName_FirstName_A1.zip**.
4. Refer to the `README` of the GitHub repository for how to generate the proper HTML report for submission.

Installation Instructions

1. If you wish to complete this assignment on Google Colab, the assignment notebook has commands that help you install necessary packages on Colab.
2. For completing this assignment locally, you need to install the base Gymnasium library, use `pip install gymnasium`. You will also need to install Jupyter Notebook, use `pip install jupyter notebook`.
3. Refer to our assignment repo for more installation instructions.

In this assignment, you will implement planning (dynamic programming) algorithms. You will use the frozen lake environment from Gymnasium (Link).

Planning

You have to use the transition probability function to solve the following set of planning problems. Refer to the helper file for details on accessing the transition probability function.

1. **Q-Value Iteration (QVI):** Implement Q-value iteration on the frozen lake environment.
   (a) What is the optimal policy and value function?
   (b) Plot $U_k = ||Q_k - Q_{k-1}||$, where Q_k is the Q-value during the k^{th} iteration.
   (c) Use the *fancy_visual* function to plot the heat maps of the optimal policy and value function.
2. **Policy Evaluation:** Consider the following polices: (i) the optimal policy obtained from QVI, and (ii) a uniformly random policy where each action is taken with equal probability. Compute the value of the these polices using:
   (a) By solving a linear systems of equations
   (b) By the iterative approach

(c) Which method is better and why?

3. **Policy Iteration (PI):** Implement policy iteration on the frozen lake environment.

   (a) What is the optimal policy and value function?

   (b) Compare the convergence of QVI and PI.