Course CS4375 – AI Techniques.
Course e-mail ai-ewi@lists.tudelft.nl
Exercise sheet Linear Inverse Reinforcement Learning
Lecturer Luciano Cavalcante Siebert, II/INSY/EEMCS

Exercise sheet: Linear inverse reinforcement Learning

*This practical exercise is **not** a deliverable for the course. The goal is to compement the lecture. If you have any questions, please contact the TAs or the lecturer (l.cavalcantesiebert@tudelft.nl).*

In this practical exercise, we will get more acquainted with inverse reinforcement learning by running some simple experiments.

First, download the content from Brightspace or from this Github repository: https://github.com/lcsiebert/IRL_assignment_1

1. *main.py*, contains the main loop. You can use it to define the parameters and run your experiments.
2. *gridworld.py*, is the environment we will be using.
3. *linear_irl.py*, constains the linear inverse reinforcement learning algorithm as described in Ng and Russel (2000)[1]

Set up
You will need to get a working python3 (either directly or via Conda) installation and install a few packages (probably you have most of them installed), namely:

- CVXOPT, a free package for convex optimization

```
pip install cvxopt
```

- Numpy

```
pip install numpy
```

- matplotlib

```
pip install matplotlib
```

Description of the environment

We will experiment with an environment called “biking in the Netherlands” (*gridworld.py*). You want to bike a given route to reach home (the upper-right grid square), departing from your initial position (lower-left grid square). You can choose to go up, down, right, or left. However, due to strong wind, your actions have a 30% chance of moving in a random direction.

Instructions

First, analyze the three files (*main.py*, *gridworld.py,* and *linear_irl.py*) and run the experiment. After that:

1) Define a new optimal policy by replacing the content of the function “optimal_policy_deterministic” in the *gridworld* file. Be creative; you can either

[1] A. Y. Ng and S. J. Russell. 2000. Algorithms for inverse reinforcement learning. In: Proceedings of the 17th International Conference on Machine Learning (ICML ’00), Stanford University, Stanford, CA, USA. https://ai.stanford.edu/~ang/papers/icml00-irl.pdf

create a function[2] or define the policy manually. However, the termination state should remain in the upper-right grid square.

2) Test different combinations of the discount factor (γ, variable: *discount*) and the penalty factor (λ, variable: *l1*), and analyze the impact on estimating the reward.

Answer and reflect on the following questions:

- **E1**: What is your new optimal policy? Please describe the reasoning behind it.
- **E2**: What was your strategy for exploring the discount and penalty factors?
- **E3**: Describe and discuss how different combinations of the discount and penalty factors impacted the estimated rewards.

[2] For example, the current function implements the following:
IF x < y:
    Go right
ELSE:
    Go left.