PRACTICAL ASSIGNMENT: INVERSE REINFORCEMENT LEARNING

Delft, September 17, 2020

This material is part of the trial lecture of Luciano Cavalcante Siebert.

1 INTRODUCTION

In this practical assignment, we will get more acquainted with inverse reinforcement learning through running some experiments with the practical example discussed in the lecture.

In the Github folder (https://github.com/lcsiebert/IRL_assignment_1) you can find the following files:

1. main.py, contains a main loop you can use to define the parameters and run you experiments.
2. gridworld.py, the environment we will be using. See section 3 for description.
3. linear_irl.py, file for the linear inverse reinforcement learning algorithm as described in Ng and Russel (2000)[1]

2 SET UP

You will need to get a working python3 (either directly or via Conda) installation and install a few packages, namely:

- CVXOPT, a free package for convex optimization

```
pip install cvxopt
```

- Numpy

```
pip install numpy
```

- matplotlib

```
pip install matplotlib
```

3 DESCRIPTION OF THE ENVIRONMENT

We will experiment with a environment called "biking in the Netherlands", that is located in gridworld.py. You want to bike a given route to reach home (the upper-right grid square), departing from your current position (lower-left grid square). You can chose to go up, right, left, down on the gridworldm but, due to a strong wind, your actions have a 30% chance of moving in a random direction.

4 INSTRUCTIONS

First of all you have to analyse the three files (main.py, gridworld.py and linear_irl.py) and run the experiments as is. Try to understand how each variable and function call works.

Your assignment will be to change the policy and predict a new reward function, without having an explicit knowledge of the "real" reward, aka the ground truth. For this, take the followings steps:

[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

1) Define a new optimal policy by replacing the content of the function "optimal_policy_deterministic" in the gridworld file. You can either create a function[2] or define the policy manually. But remember, the termination state must be in the upper-right grid square.
2) Test different combinations of the discount factor (γ, *discount*) and the penalty factor (λ, *penalty_factor*) and analyse the resulting estimated reward.

DELIVERABLE: For this assignment you are required to submit a zip-file with your code and a PDF file answering the following questions:

- Question 1: What is your new optimal policy? Please describe the reasoning behind it.
- Question 2: What was your strategy on exploring the discount and penalty factors?
- Question 3: 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.