DRL Homework 04

Leon Schmid

Deadline: July 10th 23.59

Abstract

Homework 4 asks you to implement an advanced Policy Gradient Algorithm to solve the car racing gym
https://www.gymlibrary.ml/environments/box2d/car_racing/

Contents

1 Implementing the network

Implement a network to solve this problem. Given the image-style observations in the car racing environment, I recommend using a (small!) Convolutional Neural Network. I recommend keeping it simple - a user-parameterized standard deviation will do fine! Also just assume a diagonal covariance matrix. Your network should implement the following functions:

- Given a (batch of) state(s), sample a (batch of) action(s), return the action(s) and the respective log-probability(ies)

- Given a (batch of) state(s) and action(s), return the (batch of) log probability(ies) of the action(s) given the state(s)

2 Creating trajectories

Implement a function to sample a trajectory from the environment given a model like the one implemented in Task 1. Implement functionality to either sample a full trajectory, or to sample a fixed amount of steps from a given environment.

3 'Vanilla' Policy Gradient

Implement Vanilla Policy Gradient with the network and sampling implemented above. For Policy Updates make sure you combine samples from multiple trajectories (either by unrolling multiple trajectories at once or by accumulating gradients from multiple trajectories).

4 Improving PG

Improve your PG by introducing a critic - write the respective network and training and add it to your PG implementation to create A2C! Add at least one of the following to your implementation:

- Generalized Advantage Estimation

- Change the training algorithm to implement PPO!

5 Task 04 Review

Again, find 3 other groups for which you can review homework 03. Include the written review in the folder you submit.