Reinforcement Learning 2023, Master CS, Leiden University
Assignment 2 on Deep Q Learning (DQN)

Tom Stein (s3780120) [1] **Tom Stein (s3780120)** [1] **Tom Stein (s3780120)** [1]

Abstract

This document provides a basic paper template and submission guidelines. Abstracts must be a single paragraph, ideally between 4–6 sentences long. Gross violations will trigger corrections at the camera-ready phase.

1. Introduction

In this assignment, an agent has to learn how to balance a pole in the vertical position. The environment presented is a well known and defined physic problem, being a reverse pendulum. The vertical position is correspond to the unstable equilibrium point and the pole itself is attached through a join to a cart. The goal of this learning task is to keep upright the pole while moving the cart left or right. The possible action space is made up by a set of only two possible movements $0, 1$, where 0: the cart is pushed to the left; 1: the cart is pushed to the right. The state space is composed by four values: position and velocity of the cart, angle and angular velocity of the pendulum. The agent receives a reward of $+1$ for every action performed resulted in keeping the pole into a equilibrium state. The environment is reset to the initial condition every time that the value of the angle between the pole and the vertical line is bigger than 15 or when the cart leaves the range $(-2, 4; 2, 4)$.
To tackle this problem tabular methods are not sufficient anymore, since keeping in memory all the possible states is not feasible anymore. Therefore, actions cannot be performed anymore on the basis of the Q-value stored in the table, but we need a way to generalize to unseen states. This can be achieved through the application of an Artificial Neural Network. In this assignment our deep neural network taken as input a state returns the Q values for the actions.
As a baseline comparison, we will use a random policy which has a reward around 22.
The structure of the following chapters is as follows. In section 2, 3 and 4, for each of the DQN algorithm hyperparameters optimization will be analyzed, highlighting the optimal configuration. Section 2 covers the basic case of DQN algorithm while, in section 3 the replay buffer is present and in section 4 the target network is considered. After all this architecture has been presented, section 5 is dedicate the comparison and analysis of the different DQN approaches showing their pros and their limitations.

[*]Equal contribution [1]Faculty of Science, Leiden University, Leiden, The Netherlands. Correspondence to: Tom Stein <tom.stein@tu-dortmund.de>.

References

Plaat, A. *Deep Reinforcement Learning*. Springer, 2022. ISBN 978-981-19-0637-4. doi: 10.1007/978-981-19-0638-1. URL https://doi.org/10.1007/978-981-19-0638-1.

A. Hyperparameter Scan Results