The Effect of Neuron Selection on Policy Generalization

Timothy J. Whitaker
Department of Computer Science
Colorado State University
Fort Collins, CO 80525
timothy.whitaker@colostate.edu

Abstract

The abstract paragraph should be indented ½ inch (3 picas) on both the left- and right-hand margins. Use 10 point type, with a vertical spacing (leading) of 11 points. The word **Abstract** must be centered, bold, and in point size 12. Two line spaces precede the abstract. The abstract must be limited to one paragraph.

1 Introduction

2 Methods

We experiment with 7 different state of the art reinforcement learning algorithms. These are: Advantage Actor Critic (A2C), Actor Critic using Kronecker Factored Trust Region (ACKTR), Deep Deterministic Policy Gradient (DDPG), Proximal Policy Optimization (PPO), Soft Actor Critic (SAC), Twin Delayed Deep Deterministic Policy Gradient (TD3), and Trust Region Policy Optimization (TRPO). Model implementations are provided by the open source library Stable Baselines[https://github.com/hill-a/stable-baselines].

2.1 Training

All models use a fully connected multilayer perceptron with two hidden layers of 64 neurons each for the policy architecture.

3 Results

3.1 Markovian Tasks

3.1.1 Single Pole Cart Balancing

3.1.2 Double Pole Cart Balancing

3.2 Non-Markovian Tasks

3.2.1 Single Pole Cart Balancing

3.2.2 Double Pole Cart Balancing

References

Preprint. Under review.