Assignment 2: Policy Gradients

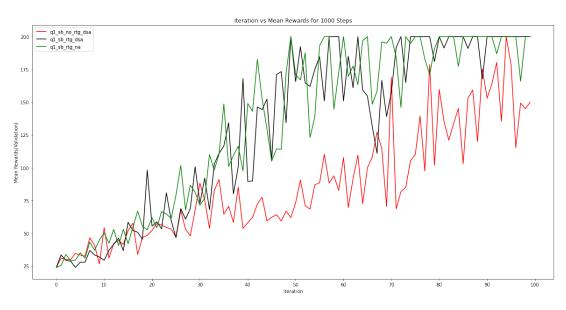
huseyinabanox@gmail.com

January 2023

1 Small-Scale Experiments

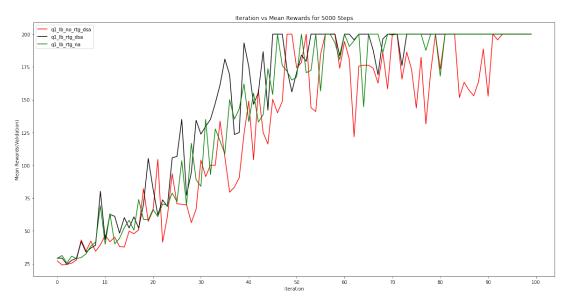
1.1 Experiment 1

PG - CartPole



Network with 2 Hidden Layers of size 64, 1K training steps per iteration, LR=0.001

Reward-to-go estimator performs better. Advantage standardization difference is not statistically significant.



Network with 2 Hidden Layers of size 64, 5K training steps per iteration, LR=0.001

Reward-to-go estimator performs better. Advantage standardization difference is not statistically significant. Increasing batch size helps the algorithm converge with fewer iterations.