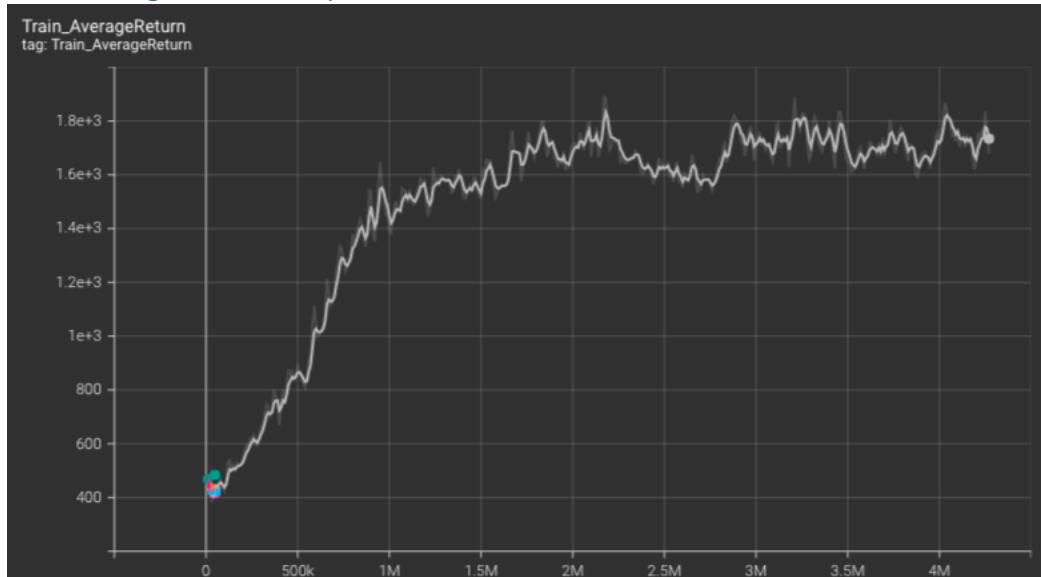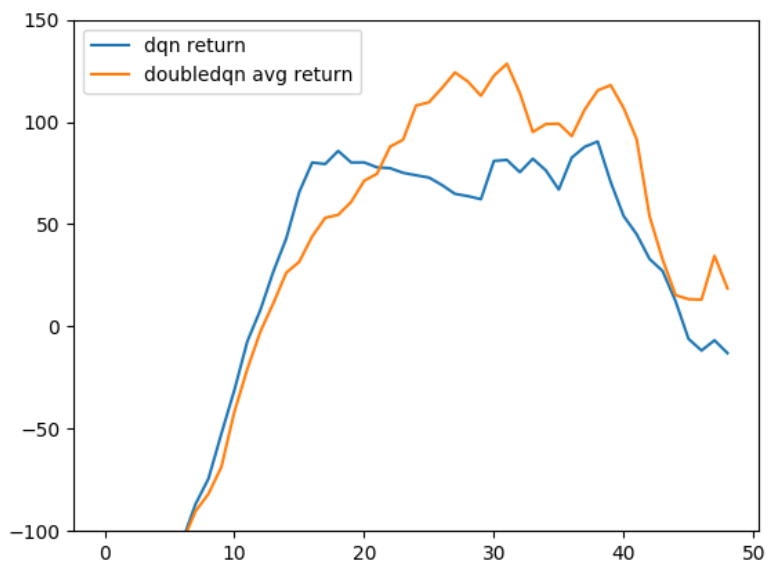# Homework #3 Report

Alejandro Escontrela
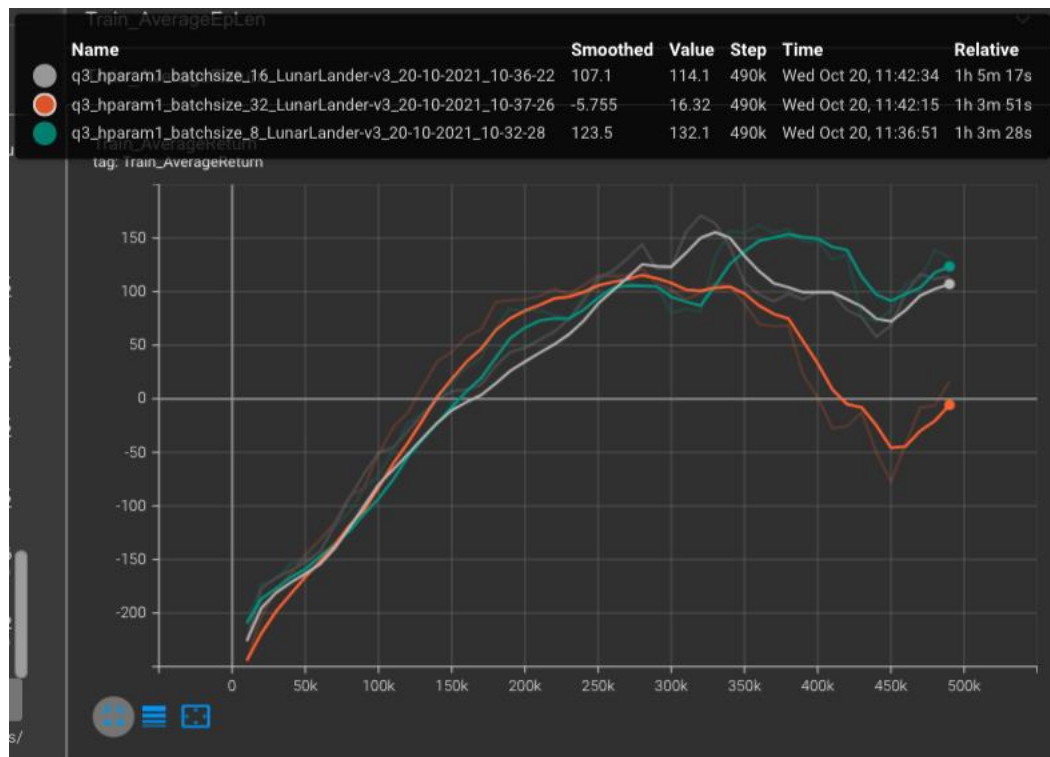
Q1: Average return on pacman

Q2: Double Q-Learning (DDQN)

Q3: DQN Hyperparameter Sweep

Name	Smoothed	Value	Step	Time	Relative
● q3_hparam1_batchsize_16_LunarLander-v3_20-10-2021_10-36-22	107.1	114.1	490k	Wed Oct 20, 11:42:34	1h 5m 17s
● q3_hparam1_batchsize_32_LunarLander-v3_20-10-2021_10-37-26	-5.755	16.32	490k	Wed Oct 20, 11:42:15	1h 3m 51s
● q3_hparam1_batchsize_8_LunarLander-v3_20-10-2021_10-32-28	123.5	132.1	490k	Wed Oct 20, 11:36:51	1h 3m 28s

Train_AverageReturn
tag: Train_AverageReturn
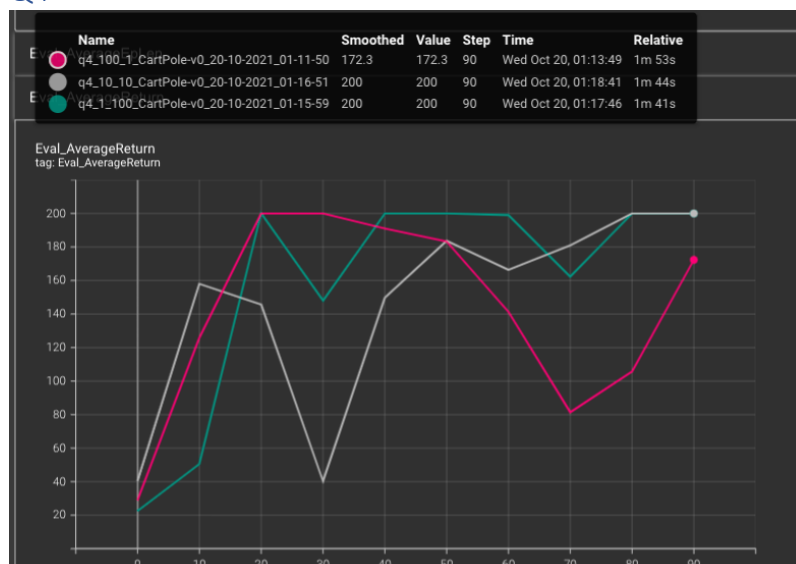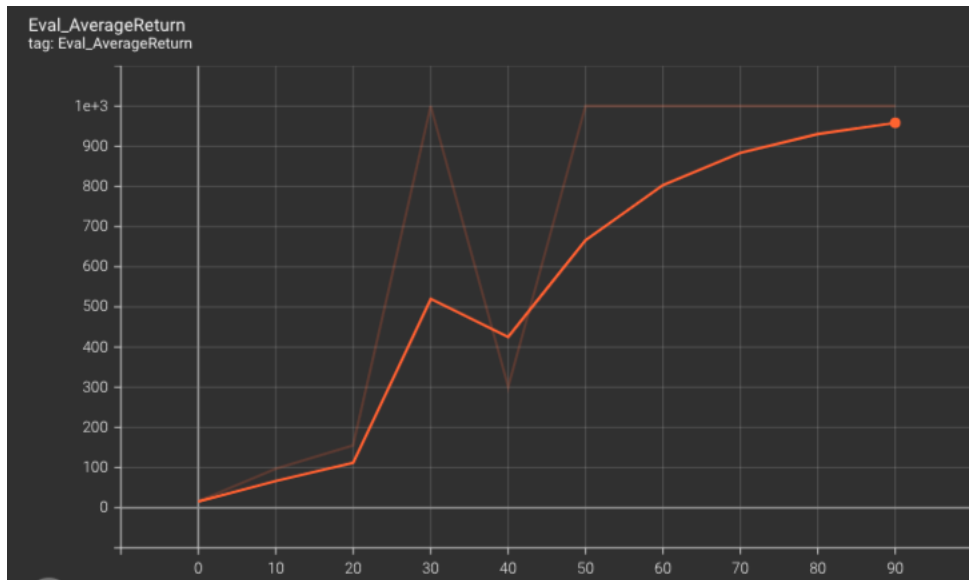

For this part of the assignment, I performed a sweep over the batchsize. Using a smaller batchsize leads to more stable learning.

Q4

Name	Smoothed	Value	Step	Time	Relative
● q4_100_1_CartPole-v0_20-10-2021_01-11-50	172.3	172.3	90	Wed Oct 20, 01:13:49	1m 53s
● q4_10_10_CartPole-v0_20-10-2021_01-16-51	200	200	90	Wed Oct 20, 01:18:41	1m 44s
● q4_1_100_CartPole-v0_20-10-2021_01-15-59	200	200	90	Wed Oct 20, 01:17:46	1m 41s

Eval_AverageReturn
tag: Eval_AverageReturn

While all three configurations achieve the highest score, it is evident that using 100 gradient steps and one critic update per update leads to more stable performance.

Q5 – Inverted Pendulum

Eval_AverageReturn
tag: Eval_AverageReturn

Q5 – HalfCheetah

Eval_AverageReturn
tag: Eval_AverageReturn