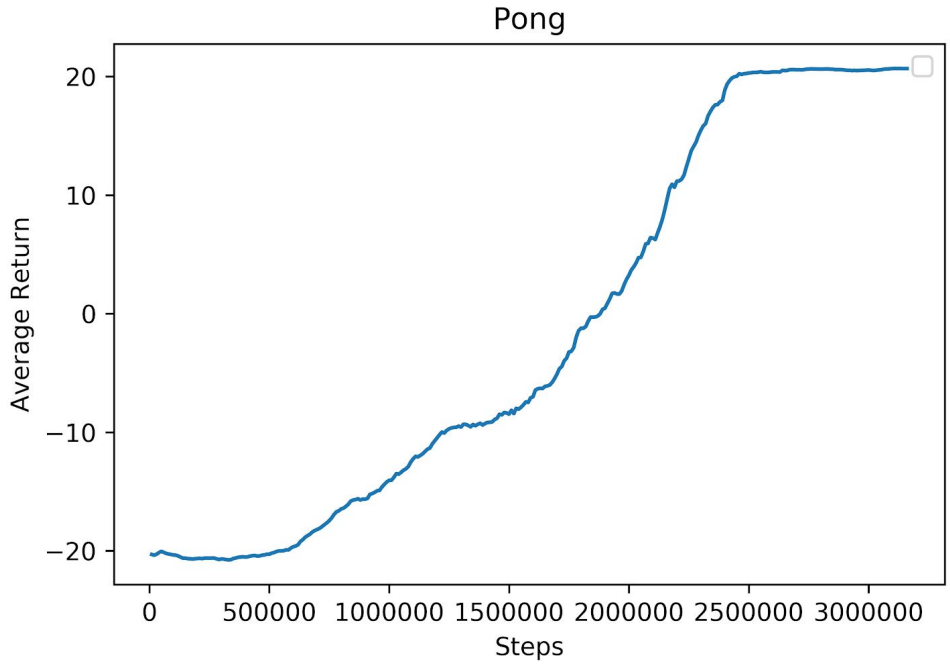
Deep RL hw2 report

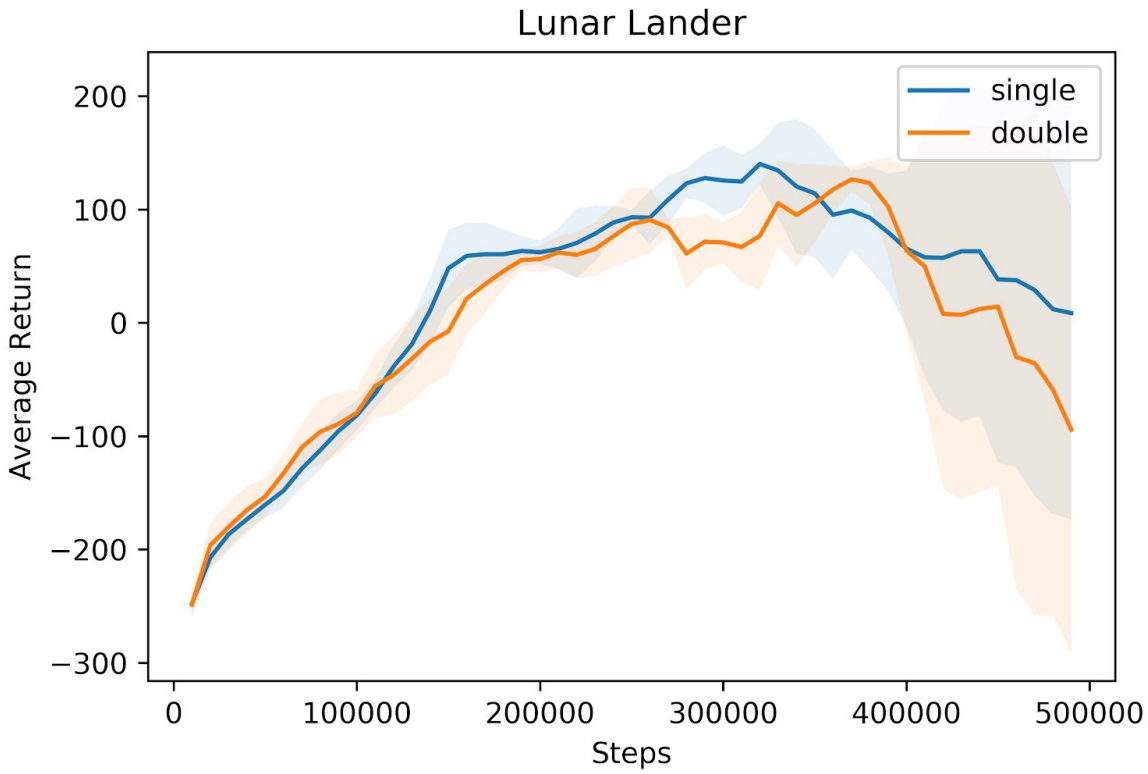
(Scripts executed exactly as they appear in the problem formulations if not stated otherwise)
(goto plotting folder to generate the plots)

Question 1

- Exploration turned off after 50000 iterations



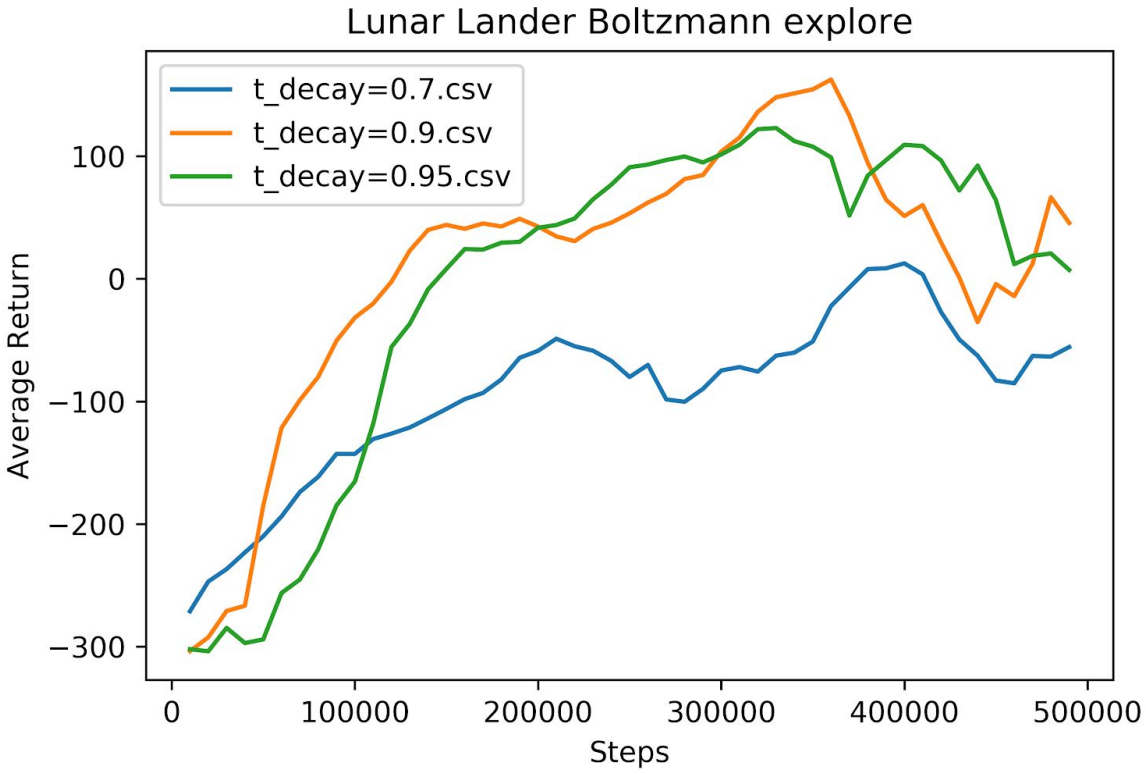
Question 2



Question 3

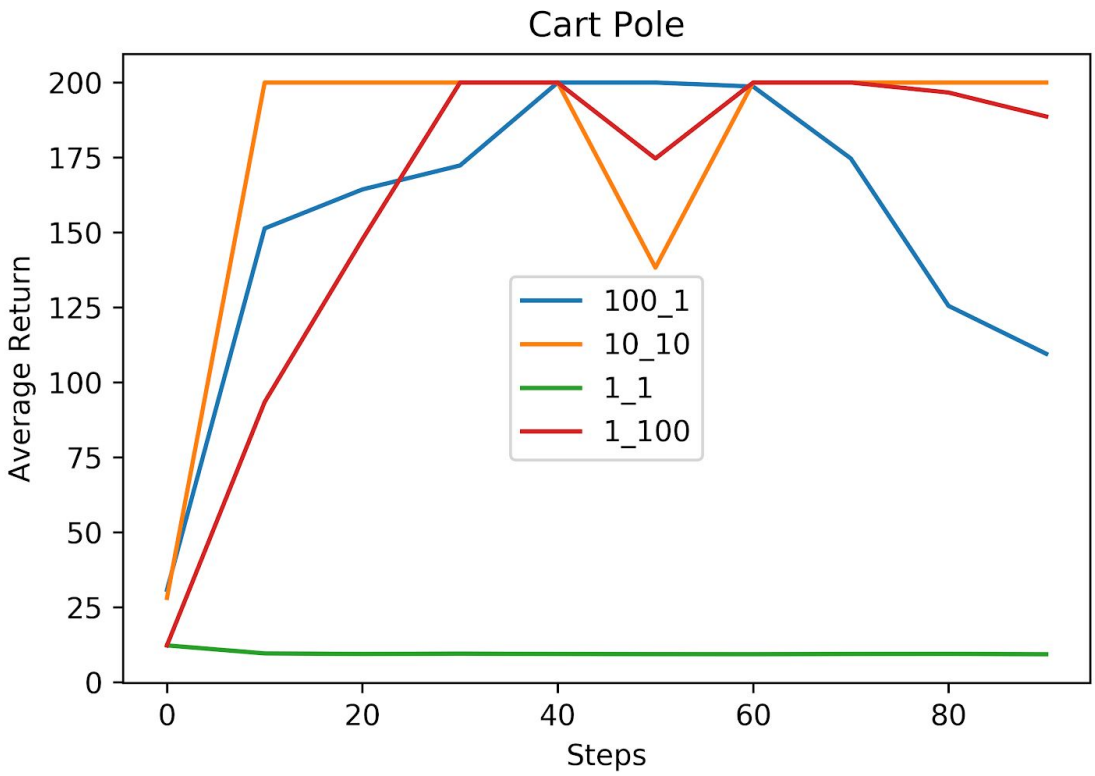
- Implemented Boltzmann exploration strategy
- Implemented Exponential Scheduler to schedule the temperature parameter of the Boltzmann exploration
    - Exponential scheduler sets the initial temperature to 1000 and multiplies the current temperature by temperature_decay every 1000 iterations
    - Bigger temperature_decay leads to more exploration
- Used LunarLander
- To run, use the template:

    python run_hw3_dqn.py --env_name LunarLander-v2 --temperature_decay <> --double_q --exp_name q3_hparam<>



Conservative behaviour leads to poor learning.

Question 4



Increasing both the number of target updates and number of gradient updates works best

Question 5 Using 10_10: