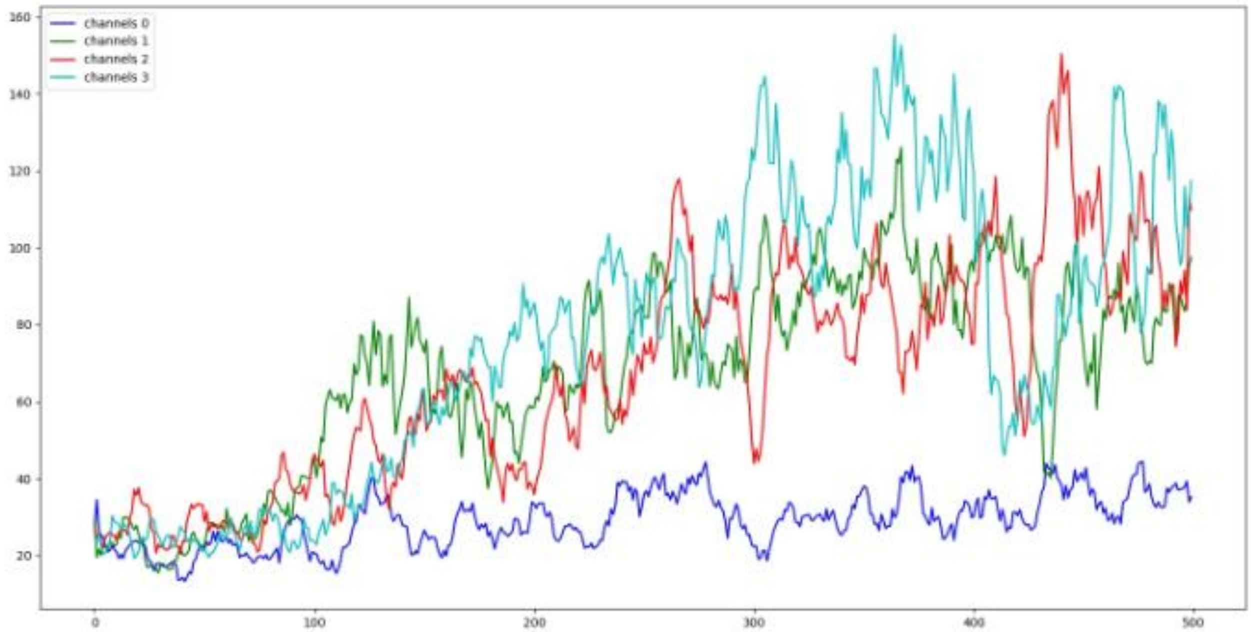
HW-2

Santhosh Vanamala - 1001964729

https://github.com/SanthoshVanamala/ActorCritic_DQN

Part 2 - Experiment I (CartPole with DQN)

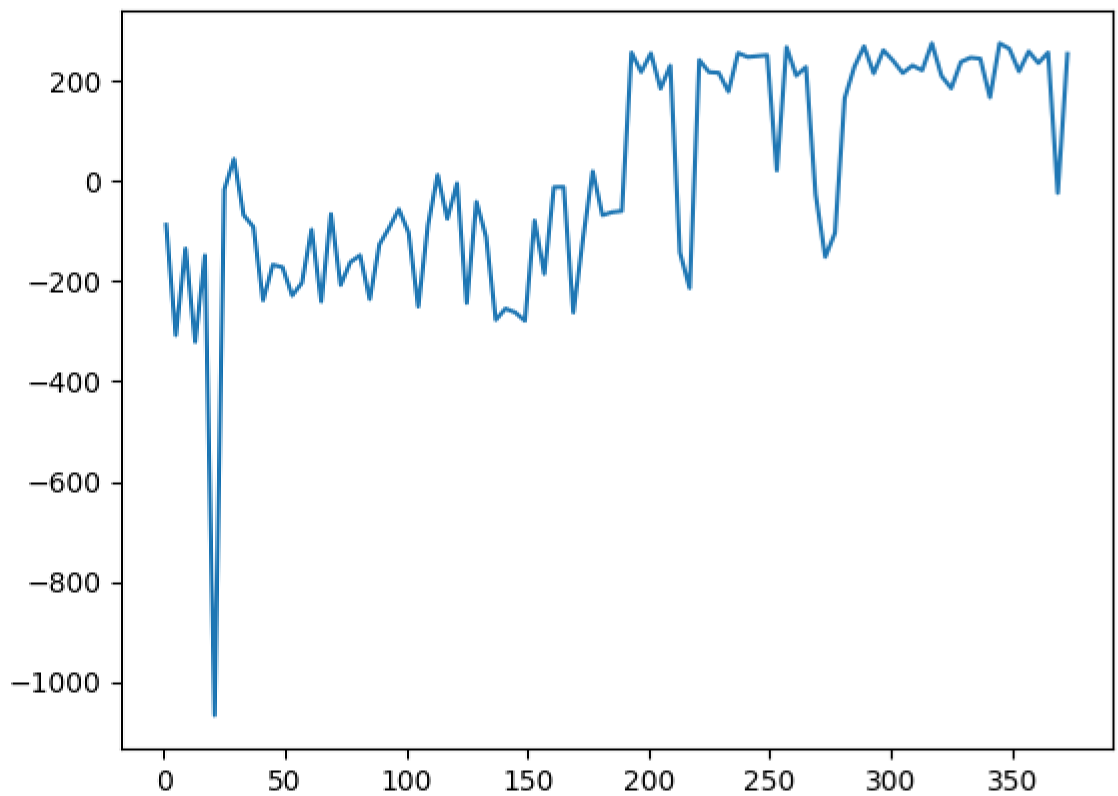


3.

a) The learning curve is influenced by the target network update rate since it determines how quickly the agent learns to generalize its Q-value estimations. The target network will be updated less often with a slower update rate, which will cause the Q-value estimations to converge more slowly. In contrast, a quicker update rate will result in more frequent updates to the target network, hastening the convergence of the Q-value estimations. However, overfitting and instability can occur when the target network is updated too frequently, which can be detrimental to learning performance. Therefore, it is necessary to adjust the target network update rate to balance stability and learning efficiency.

b) By altering the agent's behavior's balance of exploration and exploitation, changing the range for has an impact on the learning curve. A larger starting value for indicates that the agent will explore more frequently, which might speed up the process of learning the best course of action. However, excessive exploration can also lead to poor performance because the agent might not make use of its prior knowledge. On the other hand, a lower starting value for indicates that the agent will make greater use of its prior knowledge, which may result in a slower but more consistent rate of learning. In order to balance exploration and exploitation during the learning process, it is crucial to choose the ultimate value for and the pace at which it degrades. Accordingly, depending on the particular environment and task, the range for ϵ needs to be adjusted to strike a balance between exploration and exploitation.

Part 3 - Experiment II (LunarLander with AC)



2 a)Changes to the critic network update parameters, such as the number of iterations and epochs, can have a substantial effect on how well the Actor-Critic algorithm learns new material.

The critic network's precision may be increased by increasing the number of iterations and epochs, which leads to improved state-action value function approximations. This can therefore result in better actor policies and quicker convergence to the best policy. However, increasing the quantity of iterations and epochs can also lengthen training times and increase computational complexity, which can be a bottleneck for complex issues. On the other hand, cutting back on iterations and epochs might shorten training timeframes at the expense of the critic network's accuracy. This may lead to less-than-ideal policies or delayed convergence.As a result, there is a trade-off between the critic network's accuracy and computational complexity. For the Actor-Critic algorithm to function at its best, finding the ideal balance between the two is essential.