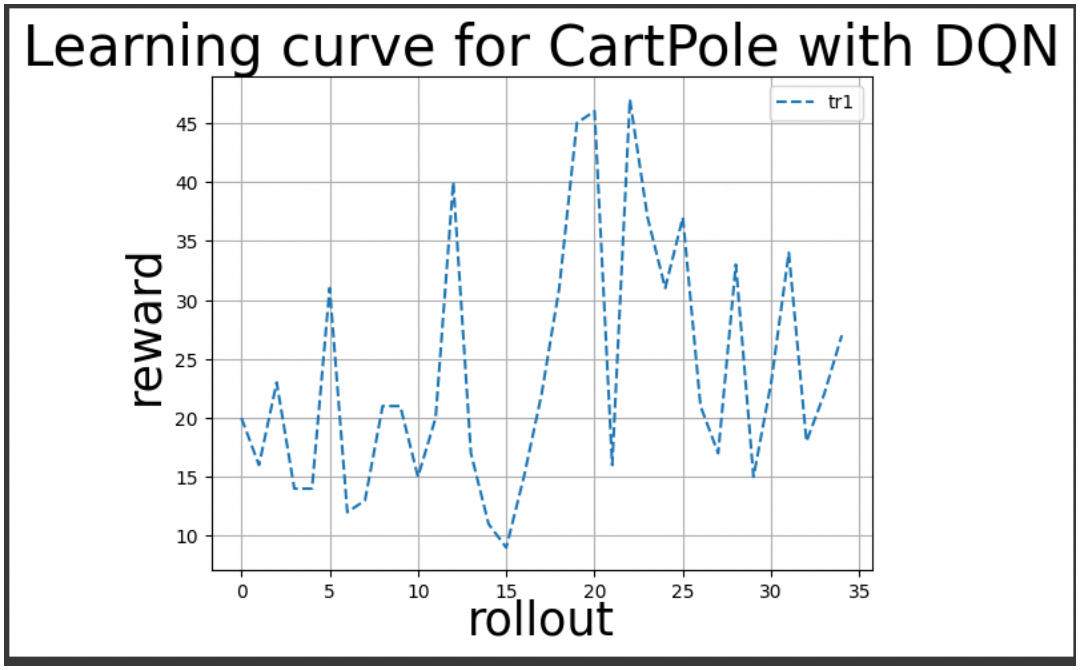
CSE 6369: SPEC TOPS ADV INTELLIGENT SYS
Assignment 02
Name: Gaurav Dilip Nale
UTA ID: 1001859699

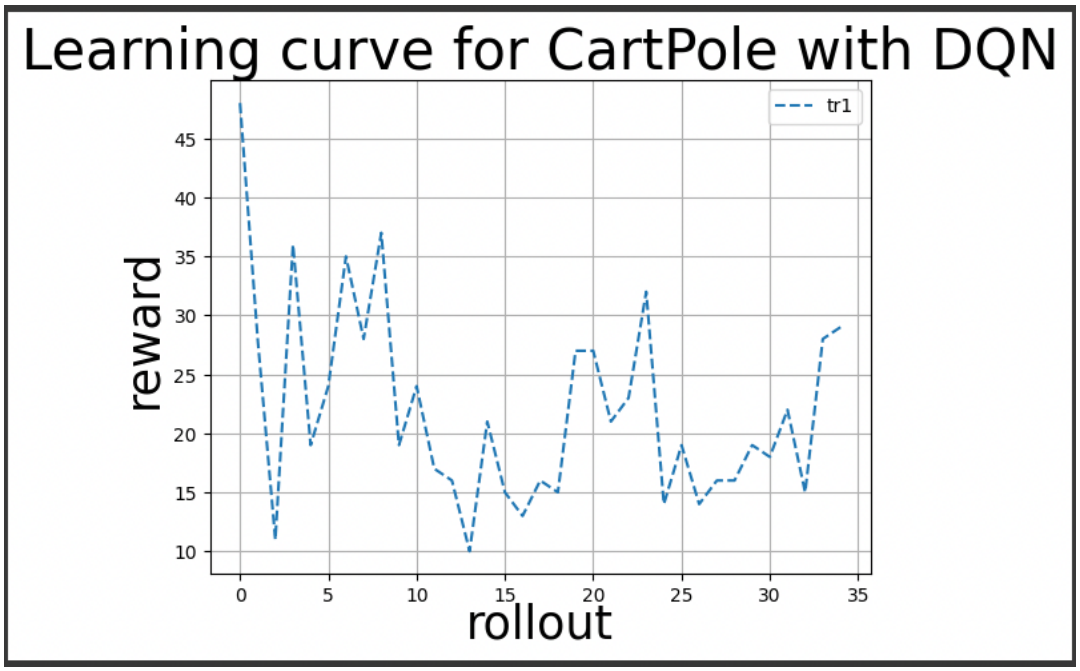
GitHub Repo: https://github.com/gauravnale/Cartpole-and-Lunar-Lander-using-Actor-Critic-and-DQN

For Cartpole using DQN:

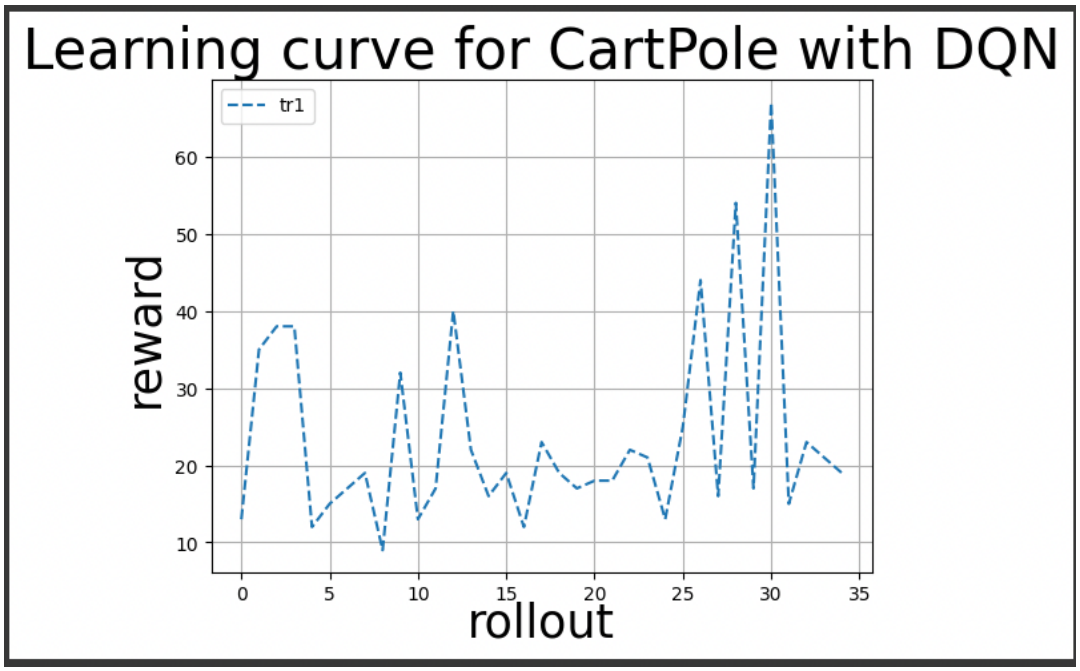
1. Trial T0: tau = 0.5



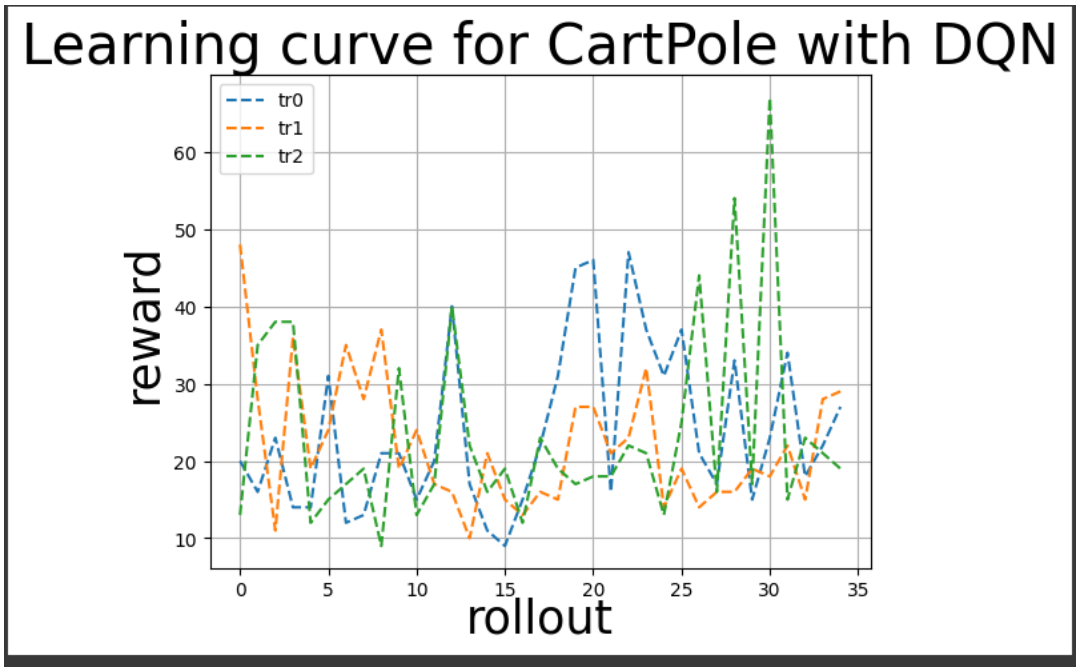
2. Trial T1: tau = 0.05



3. Trial T2: tau = 0.005

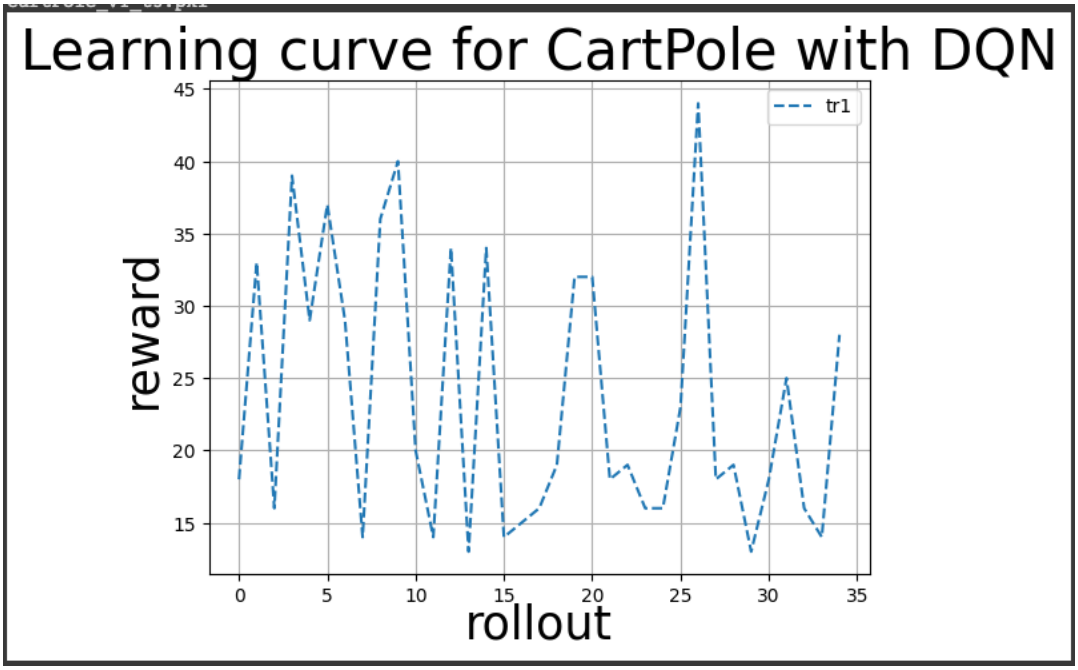


4. Graph of all three trials merged into one for comparison

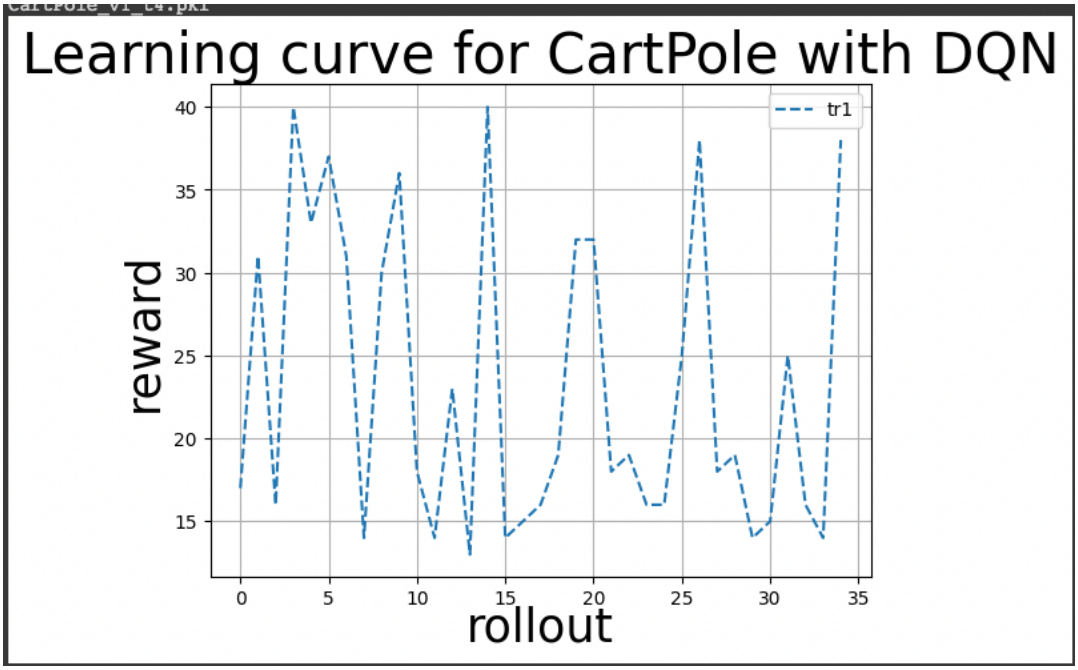


Optimum value as tau(opt):
We received highest average reward (70) for Trial 2, i.e., tau = 0.005. So, we select optimum tau value as 0.005

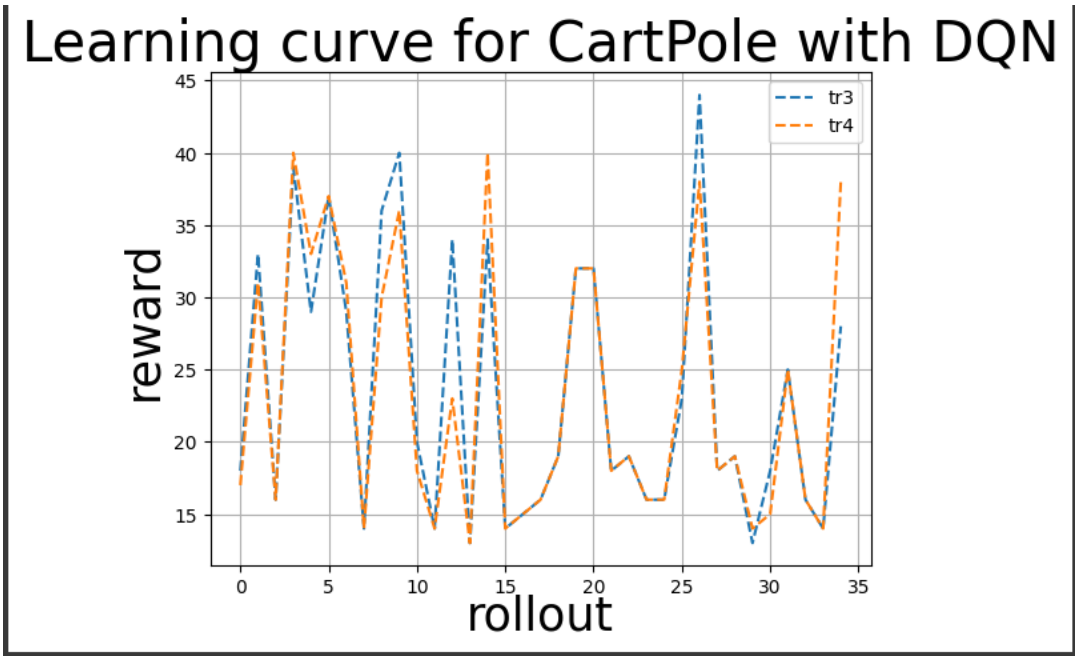
Trial 3: (Using exploration and exploitation) initial episode = 0.0, minimum episodes = 0.00



Trial 4: (Using exploration and exploitation) initial epsilon = 0.1, minimum epsilon = 0.05



Graph of all two trials (for optimum tau value) merged into one for comparison



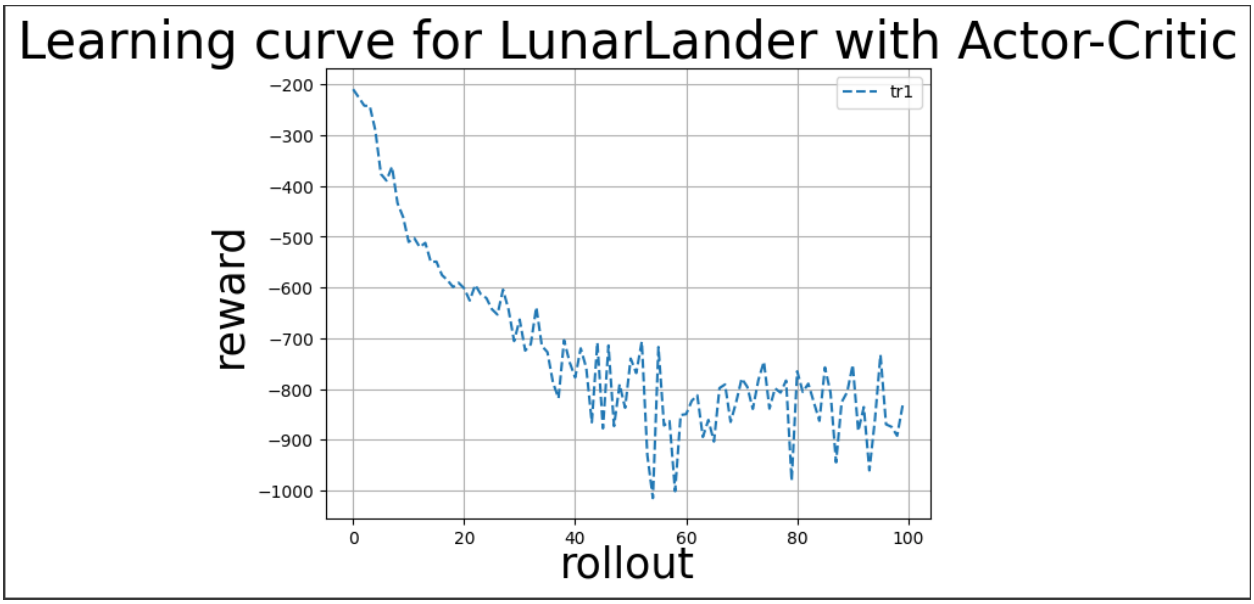
More exploration can be observed in trial 4 (initial epsilon = 0.1, minimum epsilon = 0.05). Here the learning curve is less steep and reward is obtained between fixed range.

5. How does changing target network update rate affect the learning curve? Can you justify your observation?
   - ➔ A higher value of tau means that the target network is updated less frequently and is more stable, while a lower value of tau means that the target network is updated more frequently and is more responsive to changes in the Q-network.
   - ➔ To analyze the effect of changing tau on the learning curve, we can plot the rollout vs. reward for different tau values and observe the trends in the curves. Typically, the curves with a slower update rate will show more consistent and stable progress over time, while the curves with a faster update rate will show more volatile changes and potentially more sudden improvements or setbacks.
   - ➔ From the three trials to find optimum tau value, we can see that for trial t2 there was comparatively less fluctuations in the reward value. So, there was less update for target network in this trial.

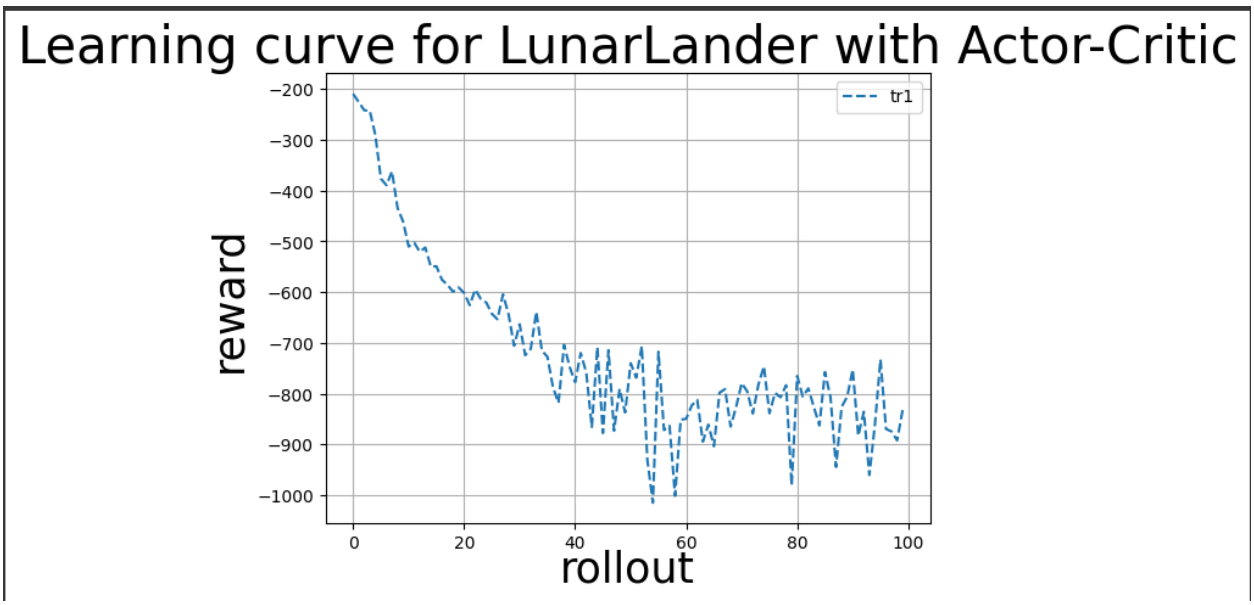
6. How does changing range for epsilon affect the learning curve? Can you justify your observation?
   - ➔ A higher value of epsilon means more exploration and a lower value means more exploitation.
   - ➔ To evaluate the effect of changing the range for epsilon on the learning curve, we can plot a graph of rollout vs reward for various epsilon values. We can expect to see that as the range for epsilon decreases, the agent starts to converge faster, and the learning curve becomes steeper. On the other hand, as the range for epsilon increases, the agent will take longer to converge, and the learning curve will be less steep.
   - ➔ From trial 3 and trial 4, we can see that learning curve is less steep in trial 4 and hence we can exploration in trial 4 and exploitation in trial 3

For Lunar Lander using Actor Critic:

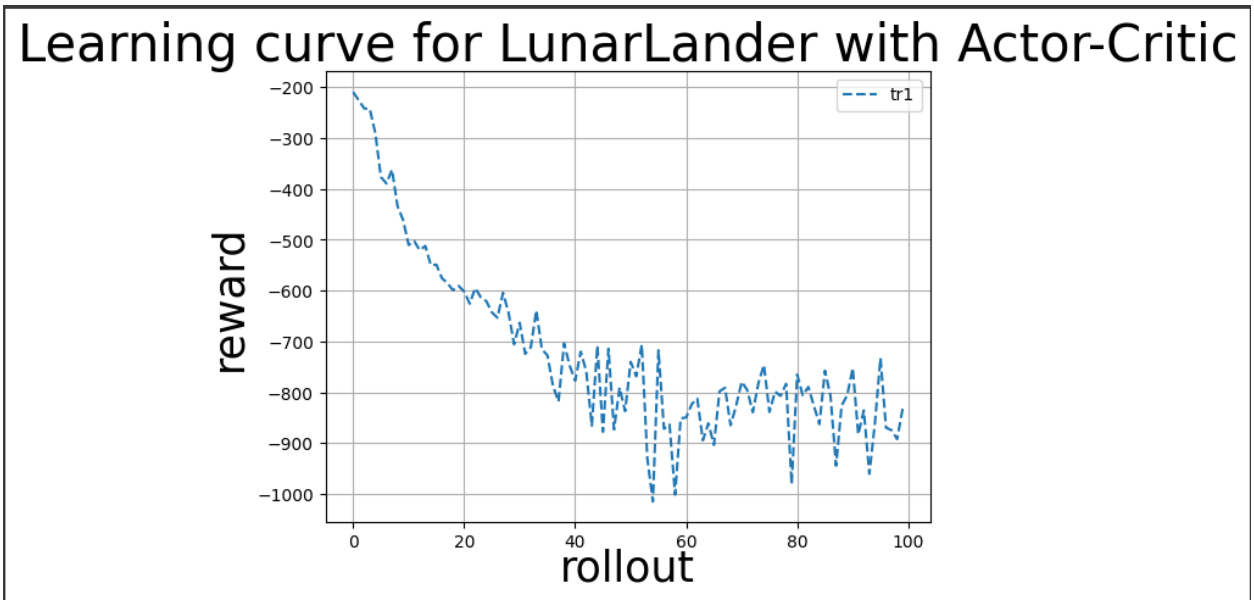
1. Trial T0: critic iteration = 1, critic epoch = 1



2. Trial T1: critic iteration = 10, critic epoch = 10



3. Trial T2: critic iteration = 20, critic epoch = 20



4. How does changing the critic network update parameters (number of iterations and number of epochs) affect learning performance? How does this justify this relationship?
   - ➔ The critic network is responsible for estimating the state-value function, which is the expected return starting from a particular state.
   - ➔ If we increase the number of iterations used to update the critic network, we allow for more updates to be made to the weights of the network, potentially resulting in a more accurate estimation of the state-value function. However, increasing the number of iterations can also increase the time required for each update, slowing down the learning process. If we decrease the number of iterations, we may speed up the learning process but also risk obtaining less accurate value estimates.
   - ➔ Similarly, if we increase the number of epochs, we allow for more updates to be made to the weights of the network for each iteration, potentially resulting in a more accurate estimation of the state-value function.