Shaira Alam
CS260R
Fall 2023
Assignment 4 Innovation Bonus
12/10/2023

1. Number of Processes (num_processes): 10 --> 20

Increasing the number of processes from 10 to 20 allows the agent to collect more experiences concurrently. When it has more processes, the agent can explore the environment in parallel so it can have more diverse samples. This increased parallelism can help in better exploration and faster convergence which will ideally improve the efficiency of the learning process.

2. Number of Steps (num_steps): 4000 --> 5000

By increasing the number of steps per update from 4000 to 5000, it allows the agent to accumulate more experience before updating its policy. This adjustment potentially capturing more complex patterns and improving the stability of the training process.

3. Number of Epochs (num_epochs): default 10 --> 30

Increasing the number of training epochs from the default 10 to 30 allows the agent to learn more from its collected experiences. Each epoch involves iterating over the entire dataset, and by performing 30 epochs, the agent has more chances to refine its policy. This change helps ensure that the agent has sufficient exposure to the data, potentially leading to more robust and effective policy updates.

4. Entropy Loss Weight (entropy_loss_weight): default 0.00 --> 0.01

Having a non-zero entropy loss weight by changing it from the default 0.00 to 0.01 encourages the agent to introduce more exploration during training. A small positive entropy loss weight helps the balance between exploration and exploitation, preventing the policy from becoming too rigid and potentially getting stuck in suboptimal solutions.