Report

Algorithm implemented: Actor-Critic

Implementation process:

1. Design the ActorCritic network by torch.
2. Design the rewards with discount factor.
3. Design the loss function (reward function) with action loss and value loss.
4. Train the model and get results within 10,000 episodes.
5. Finish training if the average reward is larger than 250.

Parameters tried:

1. Gamma (discount factor): Tried 0.5, 0.8, 0.9, and 1, found out that none of them could converge within 10,000 episodes, and thus picked 0.99 as the value of gamma through try and error.
2. Learning rate: Tried 0.1, 0.01, and 0.001, finally picked 0.005 (converge at episode 3,620) through try and error.

   (1) 0.1: cannot converge.

   (2) 0.01: converge at episode 7,040.

   (3) 0.001: need over 10,000 episodes to converge.