Deep Reinforcement Learning (CS 545)
Assignment Report
Assignment-3: Proximal Policy Optimization on Pong

Osman Furkan Kınlı – S002969

Problem definition:

In this assignment, Proximal Policy Optimization algorithm is implemented on custom Pong environment. Instead using regular Pong environment on gym, we used a custom Pong environment. The action space is discrete { up, stay, down }, and the input is not pixels as in the case of regular Pong, we decide our state space (observations). Proximal Policy Optimization algorithm is applicable to this task since we have discrete action & state spaces. The main target for this task is to score to your opponent's side by using only your paddle, and to avoid the goals for your side. We have one ball, and two players. Our opponent is just a simple AI agent which follows the ball. One score is given for one goal to each side. The match will over at a certain number of trials, and the one who has more score will win the match.

Solution approach:

To solve the Pong problem, we use Proximal Policy Optimization algorithm which give two networks for acting and criticizing. In this problem, the action space is discrete, the agent can go up or down, or decide to stay. The state space contains the position of our paddle (x, y), the position of the opponent's paddle (x, y), the position of the ball (x, y) and the angle between the ball and paddles during collision.

Observations:

{ left_paddle_x, left_paddle_y, right_paddle_x, right_paddle_y, ball_x, ball_y, ball_angle }

Reward:

if opponent wins, then punish with 1
if we win, then reward with 10
if the game still going on, punish by the distance between ball center and paddle center in order to let the agent learn to follow the ball.

Model:
Actor: 3-layer Dense network
Critic: 3-layer Dense network

Optimizer: Adam
LR: 1e-3

Results:

In our testing, our agent does not work very well. Broadly speaking, it only wins 1 game of each 5 games. We test it on 100 trials for each different run. We have 10 different runs in total.

Average result:

Simple AI 79-21 My PPO

P.S: All codes have been implemented from scratch.

Thanks to our TA Emir for giving clear intuition about discretization techniques and implementation of PPO model and custom Pong environment in his office hour.