ECE433/COS435 Introduction to RL
Assignment 8: Advanced Policy Gradient Methods & Multi-Agent Reinforcement Learning
Spring 2024

Fill me in

Your name here.

Due April 19, 2024

Collaborators

Fill me in

Please fill in the names and NetIDs of your collaborators in this section.

Instructions

Writeups should be typesetted in Latex and submitted as PDFs. You can work with whatever tool you like for the code, but **please submit the asked-for snippet and answer in the solutions box as part of your writeup. We will only be grading your write-up.** Make sure still also to attach your notebook/code with your submission.

Question 1. Prisoner's Dilemma

Game Setup: Two members of a criminal gang are arrested and imprisoned. Each prisoner is in solitary confinement with no means of communicating with the other. The prosecutors lack sufficient evidence to convict the pair on the principal charge but have enough to convict both on a lesser charge. The prosecutors offer each prisoner a bargain. The possible outcomes are:

- If A and B each betray the other, each of them serves 2 years in prison.
- If A betrays B but B remains silent, A will be set free, and B will serve 3 years in prison (and vice versa).

- If A and B both remain silent, both of them will serve only 1 year in prison (on the lesser charge).

In this case, we have the "cost" matrix (instead of a payoff matrix)[1], where the first number in each tuple is Prisoner A's years in prison, and the second number is Prisoner B's:

Prisoner A \ Prisoner B	Betray	Remain Silent
Betray	$(2, 2)$	$(0, 3)$
Remain Silent	$(3, 0)$	$(1, 1)$

Question: What is the Nash equilibrium in this game? Justify your answer and try to answer as to why it is considered a "dilemma"?

Solution

Question 2. Mixed-Strategy Nash Equilibrium (Optional)

Game Setup: Two firms, A and B, are deciding on whether to enter a market. The payoff matrix is given as follows, where the first number in each tuple represents Firm A's payoff, and the second number represents Firm B's payoff:

Firm A \ Firm B	Enter	Stay Out
Enter	$(-1, -1)$	$(2, 0)$
Stay Out	$(0, 2)$	$(0, 0)$

Question:

1. Identify any pure strategy Nash equilibria if they exist.
2. Now, we extend the pure strategy to a mixed one. Let p be the probability that Firm A enters the market, and $1 - p$ be the probability that it stays out. Similarly, let q be the probability that Firm B enters the market, and $1 - q$ is the probability that it stays out. Determine the probabilities (p and q) for which each firm should play each strategy for it to be a Nash equilibrium. (The definition of NE is still the same: neither firm can improve its expected payoff by unilaterally changing the probability distribution over its strategies.)

Question 3.

Question 3.a

First, you need to implement the policy network, which decides the actions to take, and the value network, which estimates the returns. Paste your class below:

[1]With the "cost" matrix, the agent aims to minimize costs, whereas the agent desires to maximize payoffs in standard game theory settings.

Solution

```
class PolicyNetwork(nn.Module):
    pass
```

```
class ValueNetwork(nn.Module):
    pass
```

Question 3.b

Now, you need to implement *ppo_update()*. This core function optimizes the policy and value networks using the Proximal Policy Optimization algorithm. Paste below:

Solution

```
def ppo_update(policy_net, value_net, optimizer, ppo_epochs,
    mini_batch_size, states, actions, log_probs, returns, advantages,
    clip_param=0.2):
```

Question 3.c

Finally, you should be able to run the main training. Attach the performance curves below.

Solution