ECE433/COS435 Introduction to RL
Assignment 2: Imitation Learning
Spring 2024

Fill me in

Your name here.

Due February 19, 2024

Collaborators

Fill me in

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

Instructions

The details of the assignment are provided in the accompanying .ipynb file. You should use Google Colab to run the provided .ipynb file. However, note that **we will only grade your answers within this writeup.** As such, when submitting to Gradescope, **please include the requested answer/code block in the provide solution boxes for every problem within this TeX file.** Nevertheless, make sure to still attach your notebook/code with your submission, in addition to the compiled PDF.

Problem 1. Flap Like How Flappy Sr. Taught You! (AKA Implementing Behavioral Cloning)

(a) Policy Evaluation

Paste the **entire cell** implementing policy evaluation below.

Solution

```
# YOUR CODE HERE!
```

Paste the **entire cell** for evaluating a policy that chooses actions uniformly at random below.

Solution

```
# YOUR CODE HERE!
```

(b) Defining a Policy

Paste the **entire cell** defining a policy class over discrete actions below.

Solution

```
# YOUR CODE HERE!
```

(c) Setting Up Behavioral Cloning

Paste the **entire cell** defining a behavioral cloning training loop below.

Solution

```
# YOUR CODE HERE!
```

(d) Setting Up Behavioral Cloning

Paste the **entire cell** applying behavioral cloning to `flappy_sr_notes.mat` below.

Solution

```
# YOUR CODE HERE!
```

Problem 2: Floppy the Sloppy Ruins (?) the Day (AKA An Introduction to Filtered Behavioral Cloning)

(a) What is Left???

Paste the **entire cell** applying behavioral cloning to `vandalized_notes.mat` below.

Solution

```
# YOUR CODE HERE!
```

(b) Array of Hope???

Paste the **entire cell** defining a reweighed behavioral cloning loss below.

Solution

```
# YOUR CODE HERE!
```

Paste the **entire cell** defining a training loop using the reweighed behavioral cloning loss below.

Solution

```
# YOUR CODE HERE!
```

(c) Filtering Strategy I: Trajectory-Level Reweighing???

Paste the **entire cell** implementing the trajectory-level reweighing scheme below.

Solution

```
# YOUR CODE HERE!
```

Paste the **entire cell** applying the reweighing scheme above to Filtered BC.

Solution

```
# YOUR CODE HERE!
```

(d) Filtering Strategy II: Truncated Future Return???

Paste the **entire cell** implementing the truncated future return reweighing scheme below.

Solution

```
# YOUR CODE HERE!
```

Paste the **entire cell** applying the reweighing scheme above to Filtered BC.

Solution

```
# YOUR CODE HERE!
```

Problem 3: Short-Answer Questions

For this problem, we will ask you a few questions regarding the experiments you ran to hopefully further develop your intuition for BC/Filtered BC. Limit your answer to each part to 2–3 sentences. Note that these questions are found throughout Problem 2 in the provided `.ipynb` file.

1. What does the result of the experiment in Problem 2(a) tell you about running behavioral cloning on noisy/low-quality datasets? Intuitively, why does this happen?
(Hint: Think about what the BC loss is optimizing.)

Solution

(Your answer here.)

2. Why does the trajectory-level reweighing scheme in Problem 2(c) work? How does it affect what the BC loss is doing?

Solution

(Your answer here.)

3. What is the effect of the temperature on the weighing scheme in Problem 2(c)?
(Hint: As a starting point, think about what happens to the softmax function as $\alpha \to 0$. To make it even easier to think about, consider applying the softmax to two fixed values a, b with $a > b$ as you take this limit.)

Solution

(Your answer here.)

4. One could consider a version of the reweighing scheme from Problem 2(c), where we remove the softmax and simply define the weight for τ_i as $R(\tau_i)$. While this may work in certain circumstances, can you think of some potential pitfalls of such a weighing scheme?
(Hint: Think about the values the reward function could take in all kinds of environments).

Solution

(Your answer here.)

5. Let us explore the effect of the hyperparameter T on the weighing scheme in Problem 2(d). Give a succinct description of the weights when $T = 1$. Do you expect this to work well in general? Why or why not?

Solution

(Your answer here.)

6. Can you think of a reason why one can set T in Problem 2(d) relatively small in Flappy Bird and still obtain decent performance?

Solution

(Your answer here.)