Reflections

Summary of RL Training and Changes:

Note: Agent was trained for 250 iterations, which took approximately 1 hour to train each run

Environment Setup:

- Player starts at the center of the field and is surrounded by 15 pucks
- The observation space tracks the player's position, velocity, and the positions/velocities of all of the pucks
- The action space uses discrete actions mapped to directional forces

Reward System:

+1000: reward for each puck knocked off the field
-5: penalty for unnecessary movements
-5000: penalty for going out of bounds

Training Details:

- The agent was trained using 250 iterations
- Mean, min, max, and KL divergence were logged and visualized for monitoring the learning progress

SEE: SLIDE 7 in "project 3 - group 3" powerpoint attachment

- Checkpoints were saved periodically to preserve progress and allow re-training or evaluation of earlier models (also if it timed out or wifi cut out)

Evaluation Process:

- After training, the agent's behavior was evaluated by replaying its actions using the trained policy and GIFs were created to visualize the agent's performance

SEE: SLIDE 7 in "project 3 - group 3" powerpoint attachment

Key Observations:

- The agent started with random movements hitting some pucks but gradually learned to focus on the nearest puck in bounds and move towards it

SEE: SLIDES 2 & 3 in "project 3 - group 3" powerpoint attachment

- By the end of training, it was able to clear multiple pucks in succession while staying within bounds for the most part, it will correct itself to get the pucks that have not been cleared.

SEE: SLIDE 4 in "project 3 - group 3" powerpoint attachment

Reflection on Successes and Novel Behavior:

Early Training

At the start, the player moved randomly and often didn't interact with the pucks. Some pucks moved on their own at the beginning of trials, which inflated rewards without the player actually doing anything. The player also went out of bounds frequently, leading to large penalties.

SEE: SLIDE 2 in "project 3 - group 3" powerpoint attachment

Mid Training

Around 100 iterations, the player started targeting nearby pucks but often overshot the testing area, earning penalties. Movements were inefficient, with unnecessary backtracking after pushing pucks.

SEE: SLIDE 3 in "project 3 - group 3" powerpoint attachment

Late Training

By the end of 250 iterations, the player showed a clear strategy: pushing the nearest puck off the field before moving to the next. Overshooting became rare, and movements were smoother. However, some pucks still moved without contact early in the trials.

SEE: SLIDES 4 & 6 in "project 3 - group 3" powerpoint attachment

Failures and Challenges
Early random behavior made learning slow.
Pucks moving on their own at the start gave unearned rewards.
Overshooting the field delayed progress and caused penalties.

SEE: SLIDE 5 in "project 3 - group 3" powerpoint attachment

Reflection on Key Changes:

Most Impactful Changes

1. Reward Adjustments:

Adding a penalty for idle movement encouraged the player to stay active and explore the field. A large penalty for going out of bounds (-5000) also reduced wasted effort and kept the player within the field.

2. Simplified Observation Space:

Instead of tracking all puck positions and velocities, focusing on the nearest puck made learning faster

3. Batch Size and Iterations:

Using a larger batch size (256×8) helped stabilize training, even with only 250 iterations

What We Would Do Differently:

1. Fix Environmental Issues:

Address the issue where some pucks move at the start of the trial without being touched, as it skews rewards and evaluation.

2. Explore Continuous Action Space:

Switching to continuous actions could allow for smoother and more precise movements, potentially leading to faster learning.

3. Extend Training:

Run for more than 250 iterations if computational resources allow and time allow

Most Successful Example:

Training Conditions That Led to Success:
   - Same as reflection question 1

SEE: SLIDES 6 & 7 in "project 3 - group 3" powerpoint attachment