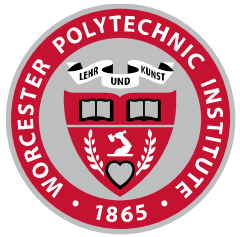


WPI

RBE 595-ST Reinforcement Learning Final Project - Lunar Lander DQN

Paul Crann

Problem Statement

- Use Deep Q Net (DQN) to train a policy capable of controlling the lunar lander agent
  - Environment: Box2d
  - Agent: LunarLander-V2
- Action Space: [do nothing, fire left, fire main, fire right]
- Observation Space: [x, y, xd, yd, theta, thetad, left contact, right contact]
- Reward Function: distance from landing pad, landers speed, lander tilt, leg contact, control effort, and crash or safe landing
- Starting State: Spawns lander at center top of environment with random force applied
- Termination State: Crash, outside of view, or sleeps

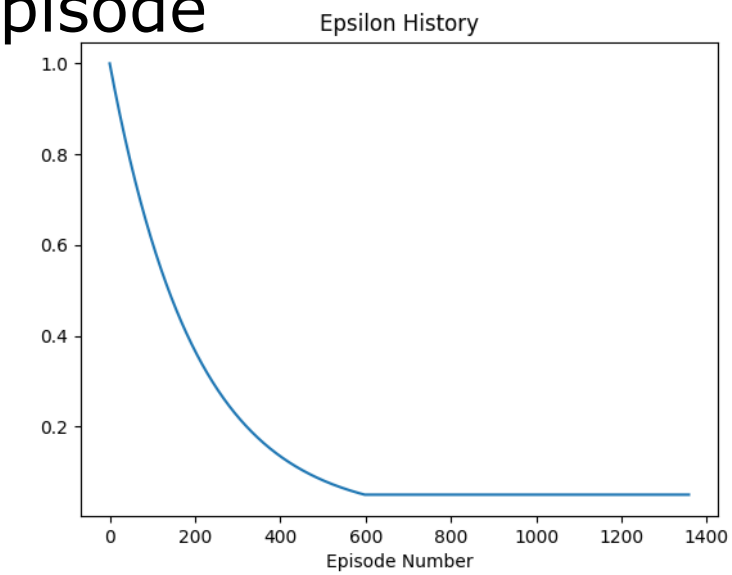
Architecture

- Using a Deep Q Net
- Action-Value Approximation: Neural Net with 3 fully connected hidden layers (64 nodes)
- Experience Replay: Stores state, action, reward, next state to then selected at random for training to increase stability
- HuberLoss vs MSELoss
  - Some outliers with very low scores -> MSE loss places a larger wight on these than the Huber loss
- Decaying Epsilon value – 1 -> 0.1 at a rate of 0.995
  - Prioritize exploration more at beginning of training, and exploitation once an efficient policy is established

```
self.fc = nn.Sequential(
nn.Linear(states, 64),
nn.ReLU(),
nn.Linear(64, 64),
nn.ReLU(),
nn.Linear(64, 64),
nn.ReLU(),
nn.Linear(64, actions)
)
```

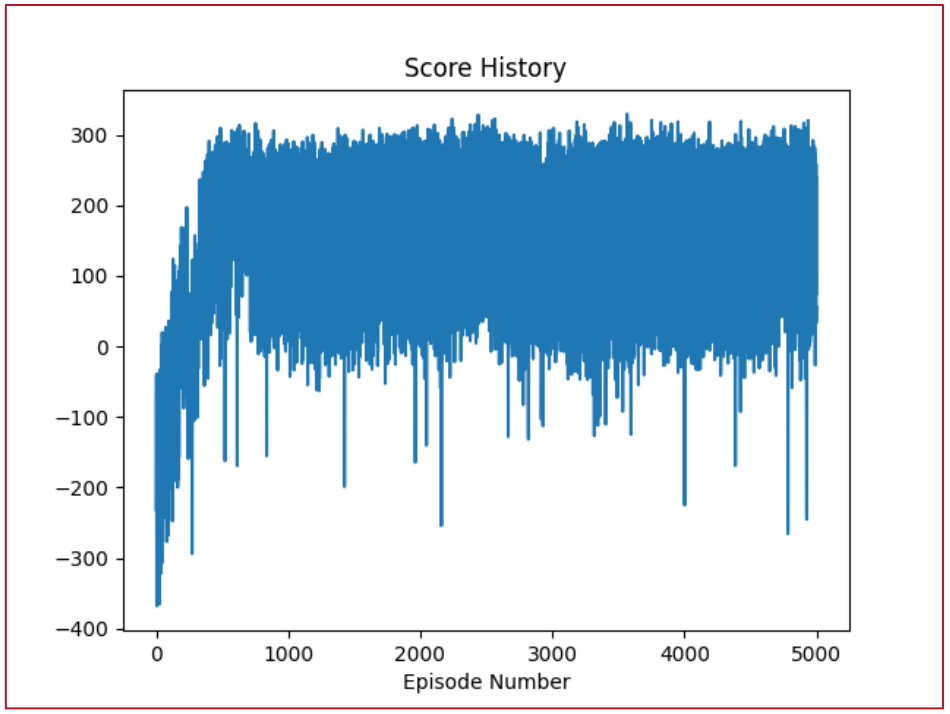
Final Hyper Parameters

- Batch Size = 128
- Learning Rate = 0.001
- Training Episodes = 5000
- Gamma = 0.99
- Epsilon = 1.0 -> 0.1 decaying at 0.995 per episode
- Learn Step = 7
- Tau = 0.002

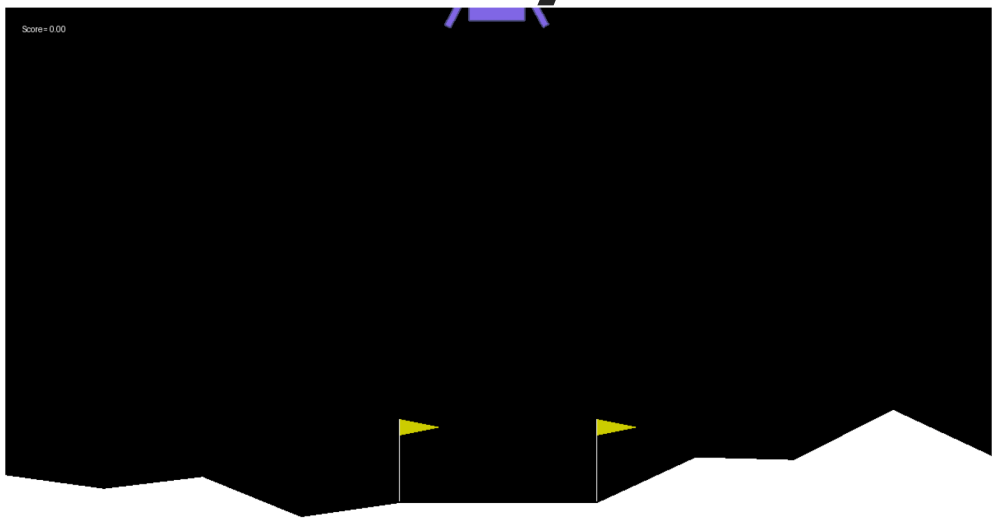


Results

- Successfully trained a policy to control the lunar landing agent
- Average score of final model over 100 runs = 249.3

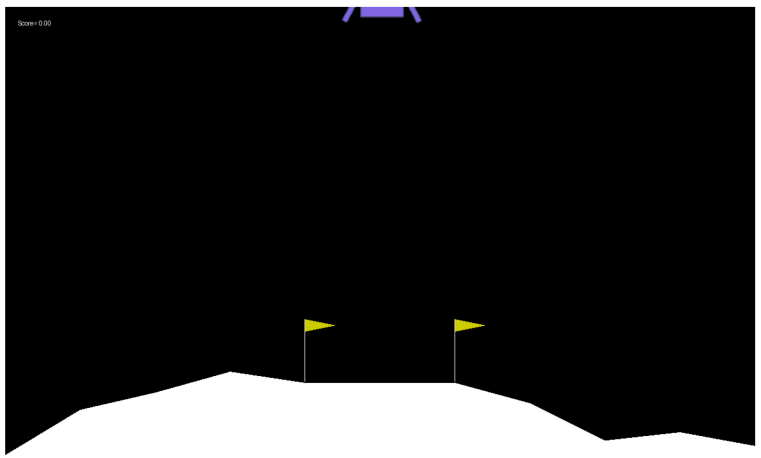


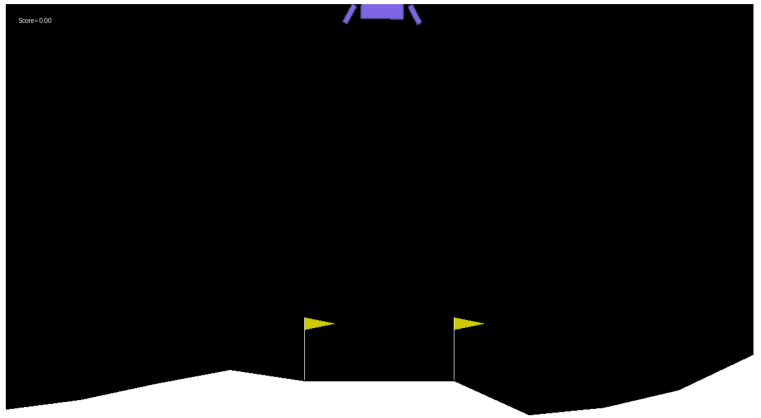
Final Policy Trials



Challenges

- Biggest Challenge was tunning model hyper parameters
  - Ex) Gamma = 0.95 vs 0.99





Future Work

- Future model tunning to create a better policy
- Implement a Double Deep Q Net
- More challenging environment with substantial horizontal distance to landing zone