Policy Gradient Methods with Pong

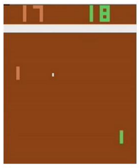
Jeffrey Hu, Gregory Soh
Department of Computer Science, Stanford University

Goals

- Apply Deep Reinforcement Learning methods towards solving Pong
- Difficulty lies in understanding image, large state space, and delayed rewards

Methods

- Used AWS EC2 p2.xlarge cluster (NVIDIA Tesla K80 GPU) for processing
- OpenAI Gym with the stochastic environment Pong-v0
- Tried vanilla policy gradient with various models



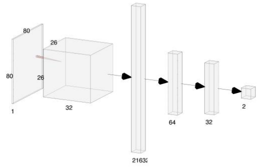
Pong-v0

Models

- Model 1: One 200 neuron hidden layer with ReLU activations
- Model 2: Convolutional layer that feeds into two fully connected layers
- Model 3: Changed discount rate from 0.99 to 0.95

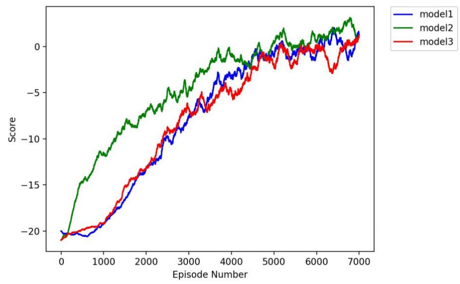
<u>Commonalities</u>

- Adam optimizer with learning rate of 0.001
- Output of two softmax classes



Model 2 Architecture

Results



- Each model was run twice and the average values for each model was plotted.
- Model 2 converges the fastest, and models 1 and 3 converge at the same rate.

- Discount factor has no noticeable impact on convergence rate (also tried with discount factor of 0.9)
- Convolutional neural network expectedly leads to faster convergence

<u>Graphs in progress:</u>

- Training with 3 actions (up, down and stay), instead of only 2 actions (up and down)

	Model 1	Model 2	Model 3
# of trials until 0	5056	4616	5129
Highest score	2.05	3.13	1.22
Mean score (last 1k)	0.36	1.66	-0.59
Mean score (overall)	-7.74	-4.52	-7.93

Future

- Implement using Q-learning and Deep Q-networks (DQN). Recent work has shown that DQN can reach human-level performance on a wide variety of Atari games.
- Implement using Asynchronous Advantage Actor-Critic (A3C). Research has shown that exploiting asynchronicity achieves better results than vanilla DQN and do so with fewer computational resources.

References

[1] James, Vincent *DeepMind's AI became a superhuman chess player in a few hours, just for fun* The Verge (2017) https://www.theverge.com/2017/12/6/16741106/deepmind-ai-chess-alphazero-shogi-go
[2] Mnih, Volodymyr et al. *Asynchronous Methods for Deep Reinforcement Learning* (2016) https://arxiv.org/pdf/1602.01783.pdf
[3] Dhruv Parthasarathy W*rite an AI to win Pong from scratch with Reinforcement Learning* Medium (2016) https://medium.com/@dhruvp/how-to-write-a-neural-network-to-play-pong-from-scratch-956b57d4f6e0
[4] Guillaume, Lample et al. *Playing FPS Games with Deep Reinforcement Learning* (2017) http://www.cs.cmu.edu/~dchaplot/papers/aaai17_fps_games.pdf
[5] Istvan, Szita *Reinforcement learning in games* (2011) https://link.springer.com/chapter/10.1007/978-3-642-27645-3_17
[6] Mnig, Volodymyr et al. *Playing Atari with Deep Reinforcement Learning* (2013) https://arxiv.org/pdf/1312.5602v1.pdf
[7] Andrej Karpathy *Deep Reinforcement Learning: Pong from Pixels* (2016) http://karpathy.github.io/2016/05/31/rl/