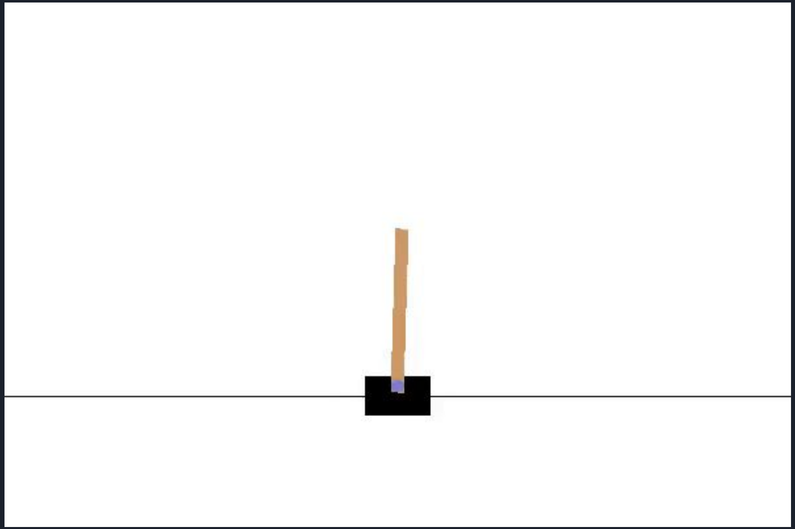
Reinforcement Learning Project 02

Group 9

Alceu Emanuel Bissoto
George Corrêa de Araújo
Jing Yang
Levy Gurgel Chaves
Stephane de Freitas Schwarz

Introduction

- CartPole-v1[1]



[1]https://gym.openai.com/envs/CartPole-v1/

REINFORCE

- Monte-Carlo Policy Gradient
- On-policy
- Actor-only
- Estimates the policy directly

Implementation from Ching -Yao Chuang[1]

[1] https://github.com/chingyaoc/pytorch-REINFORCE

Advantage Actor Critic (A2C)[1]

- On-Policy
- Actor Critic Method
  - Actor: selects action, updates the policy in the direction suggested by the Critic
  - Critic: critiques the action selected, providing feedback on how to adjust
- Concept of Advantage
- Implementation based on Dongmin Lee's code[2]

[1]https://arxiv.org/abs/1602.01783

[2]https://github.com/dongminlee94/deep_rl/blob/master/agents/a2c.py

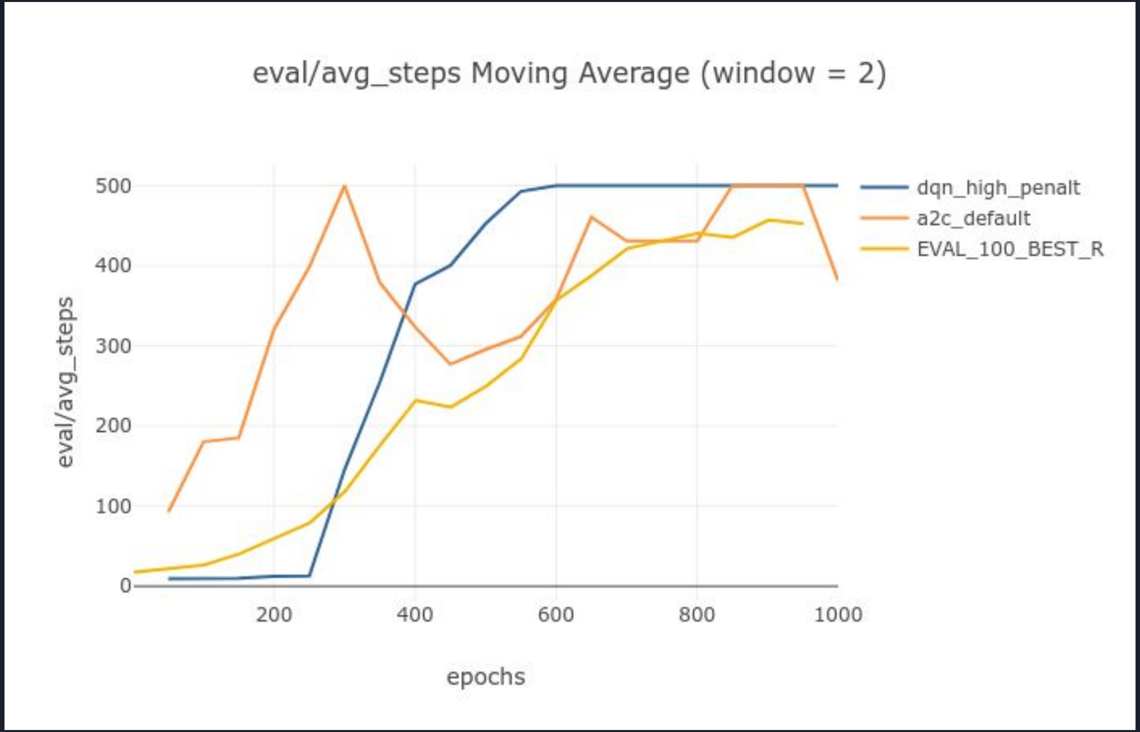
Deep Q Networks (DQN)[1]

- Off-policy
- Use neural networks to approximate the Q function
- Replay Memory plus epsilon-greedy exploration to help training

[1]https://arxiv.org/abs/1312.5602

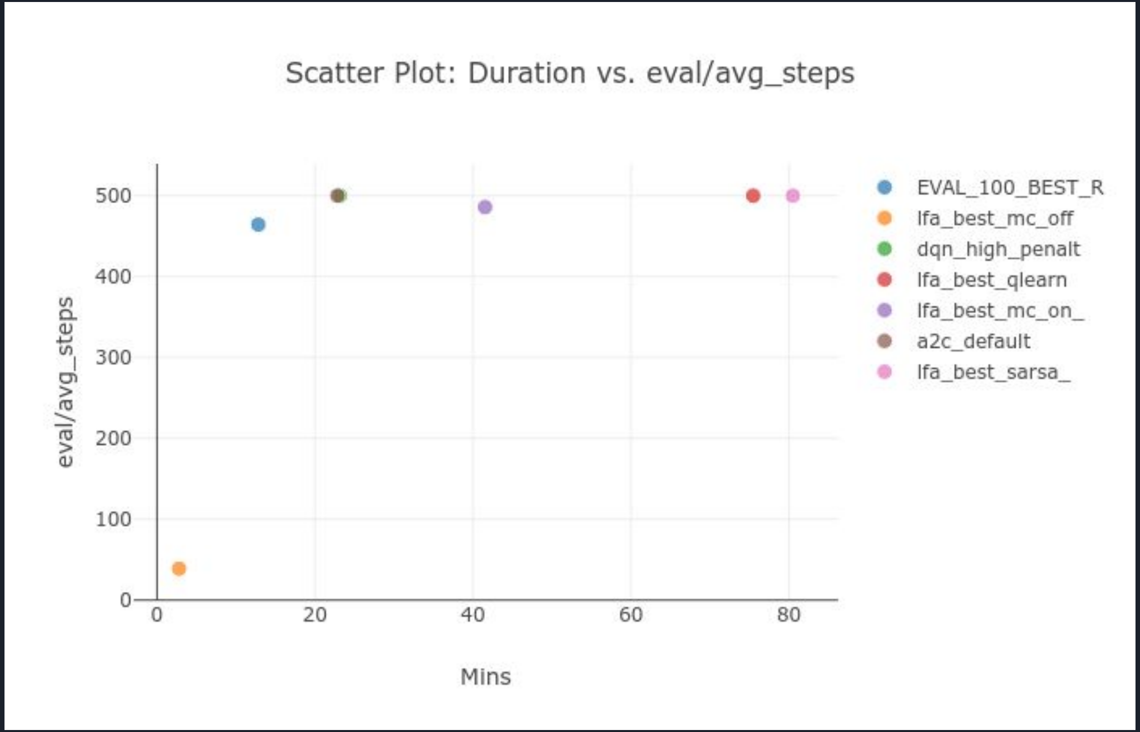
Non-Linear Function Approximators

Moving average of evaluation steps per 100 episodes



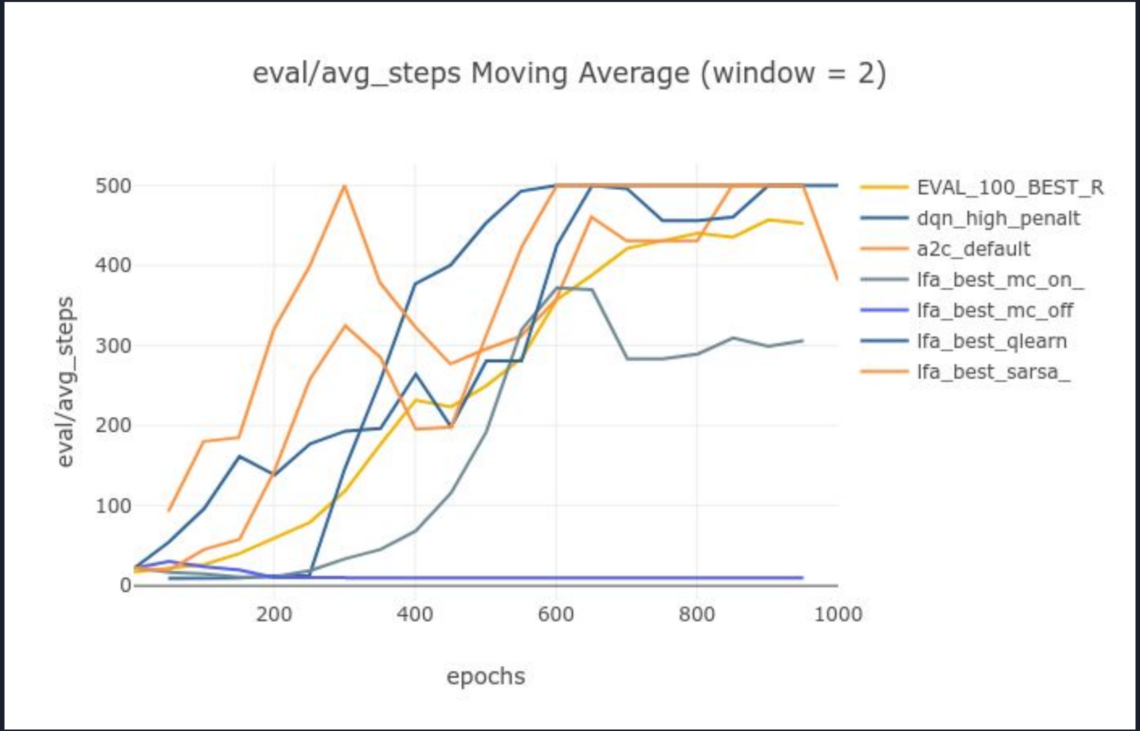
Linear X Non-Linear

Training and evaluation duration of all methods



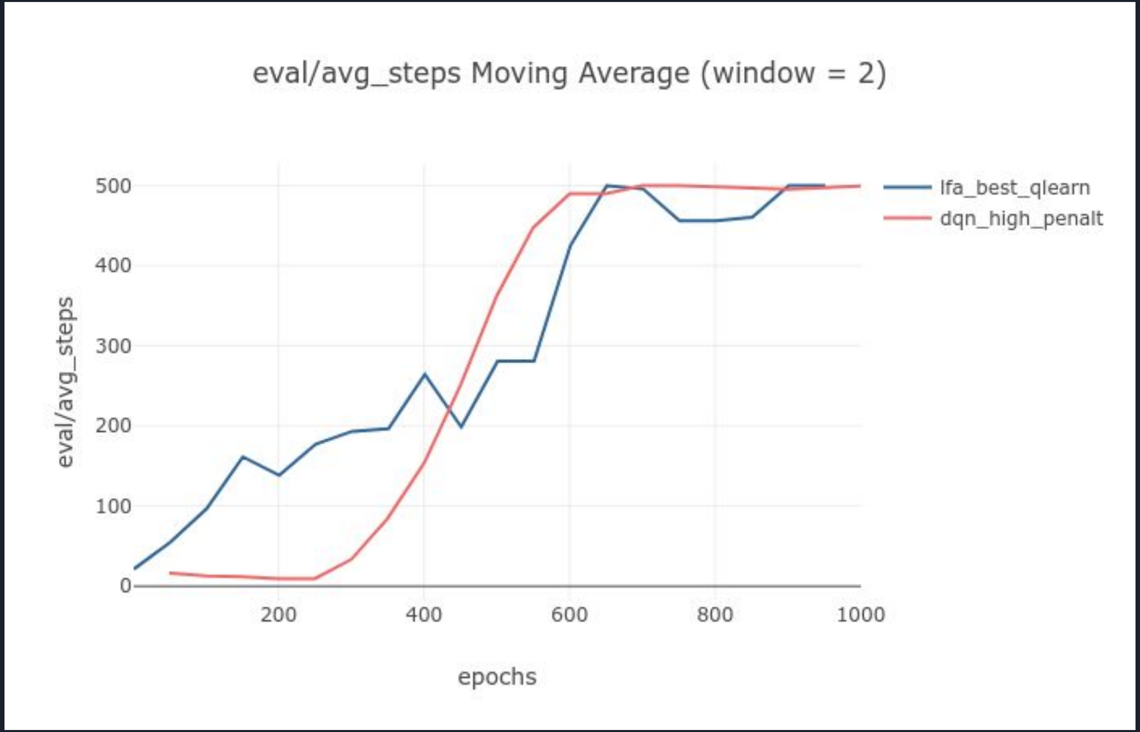
Linear X Non-Linear

Moving average of steps over 100 episodes for all methods with best hyperparameters found



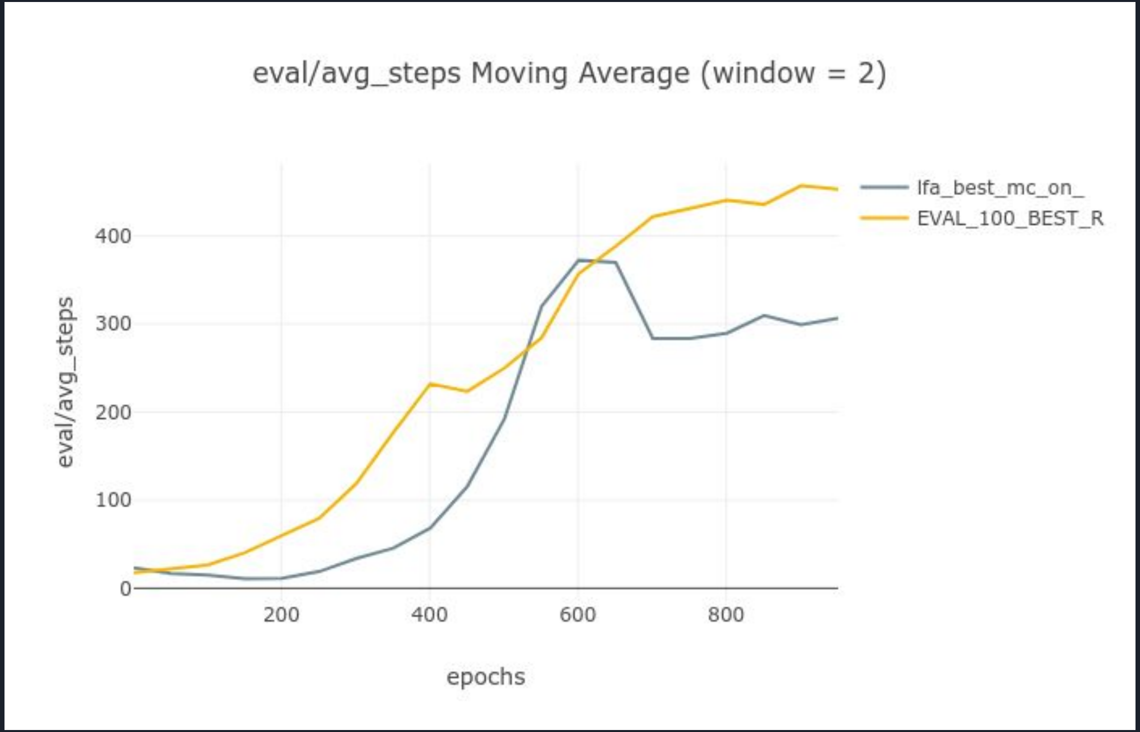
Linear X Non-Linear

LFA Q-learning X DQN



Linear X Non-Linear

LFA Monte Carlo X REINFORCE



Conclusion

- Three potential methods to solve CartPole problem
- DQN and Actor-Critic performed better
- REINFORCE is faster, but has high variance
- For simple problems, like Cartpole, LFA can perform better than deep RL

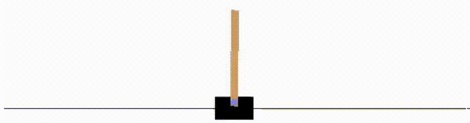
References

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T. P., Harley, T., Silver, D., Kavukcuoglu, K. Asynchronous Methods for Deep Reinforcement Learning. arXiv:1602.01783 [cs.LG].

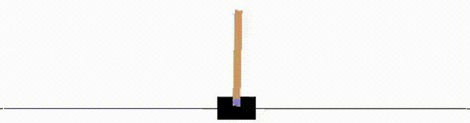
Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., Riedmiller, M. Playing Atari with Deep Reinforcement Learning. arXiv:1312.5602 [cs.LG].

Patacchiola, Massimiliano. 2018. Dissecting Reinforcement Learning-Part.8. https://mpatacchiola.github.io/blog/2018/12/28/dissecting-reinforcement-learning-8.html

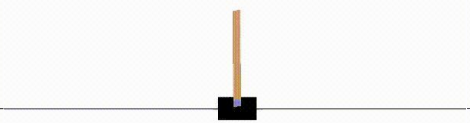
Monte Carlo



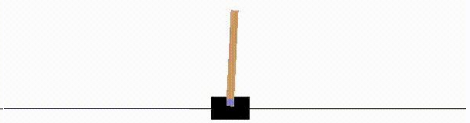
Q-Learning



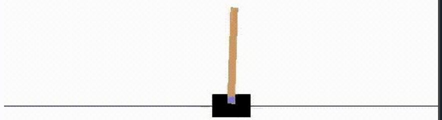
Sarsa-Lambda



Linear Approximation Sarsa-Lambda



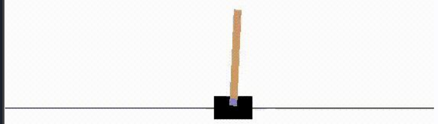
Linear Approximation MC On-policy



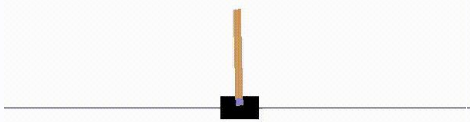
That’s all folks.

https://youtu.be/PEhwNhPwdaA

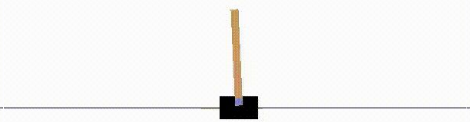
Linear Approximation MC Off-policy



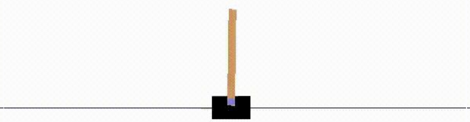
A2C



DQN



REINFORCE



Linear Approximation Q-Learning