Report

Learning Algorithm

The agents are trained using DDPG framework with shared experience.

Model and Hyper Parameters

Both Actor and Critic networks have two hidden layers of 512 and 256 nodes. Inputs are normalized using BatchNorm1d. Each layer comes with RELU activation. The final output activation is tanh.

The actor has a learning rate of 0.0001 and the critic has a learning rate of 0.0003. The agent uses the Ornstein–Uhlenbeck process to explore. The Ornstein–Uhlenbeck process has theta of 0.15 and sigma of 0.2.

Refer to the `tennis.ipynb` and `config.py` for hyper-parameters

Result

The environment is solved in under 60 episodes. At 100th episode, the average score has reached 2 and the score for each episode has consistently stayed above 3!

Future Ideas

- Experimented with different batch sizes, bigger batch sizes do not necessarily improve performance but slow down training
- Consider implementing PPO
- Consider using simple DDPG with the perspective of single player and see if the performance improves