

1) Introduction

- The goal is to get an average score of +30 over 100 consecutive episodes.

2) Environment

- Option2: Second version, 20 Agents

2) Learning Algorithm

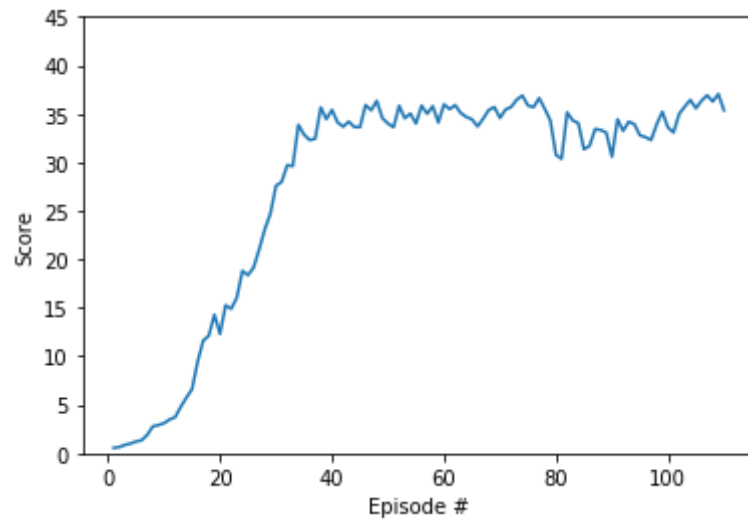
- Deep Deterministic Policy Gradients(DDPG)
- Shared Memory
- Exploration: OUNoise

3) Parameters

- Actor: 2 Hidden fully-connected layers(512, 256 units)
- Critic: 2 Hidden fully-connected layers(512, 256 units)
- Actor learning rate: $1e-4$
- Critic learning rate: $3e-4$
- Gamma: 0.99
- Soft update(τ): $1e-3$
- Memory size: $1e6$
- Batch size: 256
- Optimizer: Adam

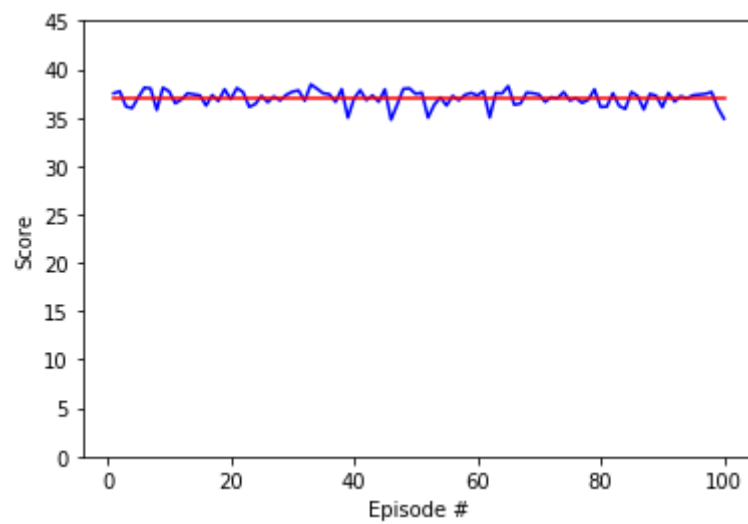
4) Training

- Training episodes: 110



5) Results

- Average score: 37.03



6) Future work

- To improve learning method, Distributed Distributional Deterministic Policy Gradients(D4PG) will be implemented