Report

Learning Algorithm

Model Design

The Actor's model consists of two hidden layers,with **400** and **300** units. Each layer comes with ReLU activation. The final output to action size uses tanh activation function to match the value

The Critic's model consists of three hidden layers with **400**, **300** and **128** units. Each comes with ReLU activation. The final layer is an output of size 1 to evaluate the state-action value.

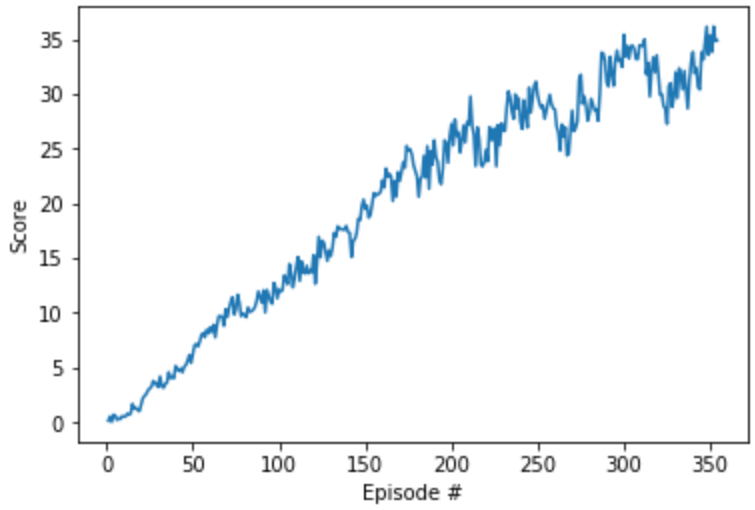
Hyperparameters

The learning agent uses a buffer for experience storage that can hold up to $\mathbf{10^6}$ timesteps of experiences. The learning agent learns from batches of **128** experiences. Soft updates use tau of **0.99.** Learning rates for the actor and the critic are $\mathbf{10^{-4}}$ and $\mathbf{10^{-3}}$.

The OU Noise process uses theta of **0.19** and sigma of **0.22**.

Scores

The agent solves the environment in approximately 350 episodes. The plot of average scores/episode is shown below.



Ideas for future work

- Run model on GPU using CUDA to speed up the training process.
- More Hyper-parameter tuning.
- Implementing a PPO model.