

UDACITY

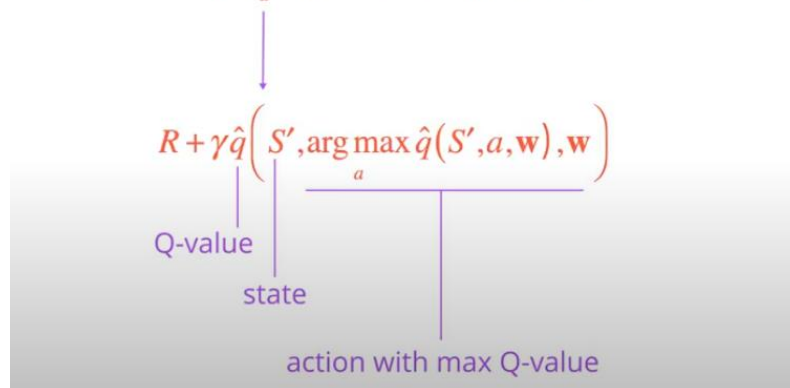
DEEP REINFORCEMENT LEARNING  
NANODEGREE UDACITY

Navigation Project

## 1. Learning Algorithm

# Overestimation of Q-values

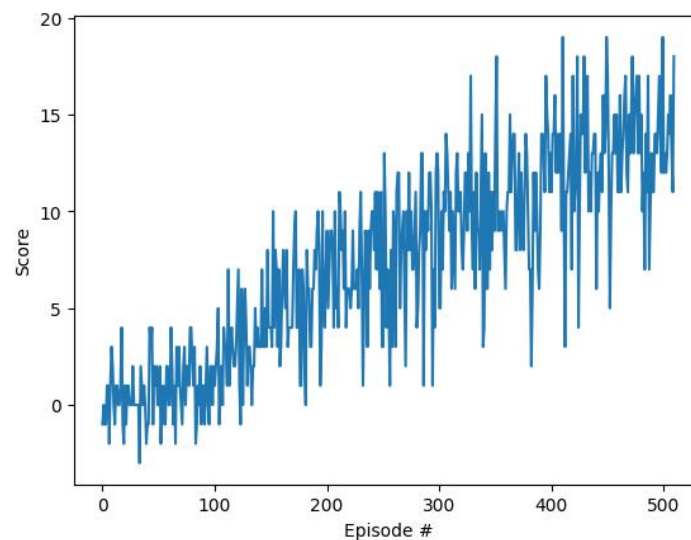
$$\Delta \mathbf{w} = \alpha \left( R + \gamma \max_a \hat{q}(S', a, \mathbf{w}) - \hat{q}(S, A, \mathbf{w}) \right) \nabla_{\mathbf{w}} \hat{q}(S, A, \mathbf{w})$$



## 2. Plot of Rewards

I needed 410 episodes to solve the environment:

Episode 100	Average Score: 0.69
Episode 200	Average Score: 4.23
Episode 300	Average Score: 7.70
Episode 400	Average Score: 10.11
Episode 500	Average Score: 12.97
Episode 510	Average Score: 13.04
Environment solved in 410 episodes!	Average Score: 13.04



## 3. Ideas for Future Work

To improve convergence speed, the developments covered in the dual DQN course can be used to help reduce overestimation of action values