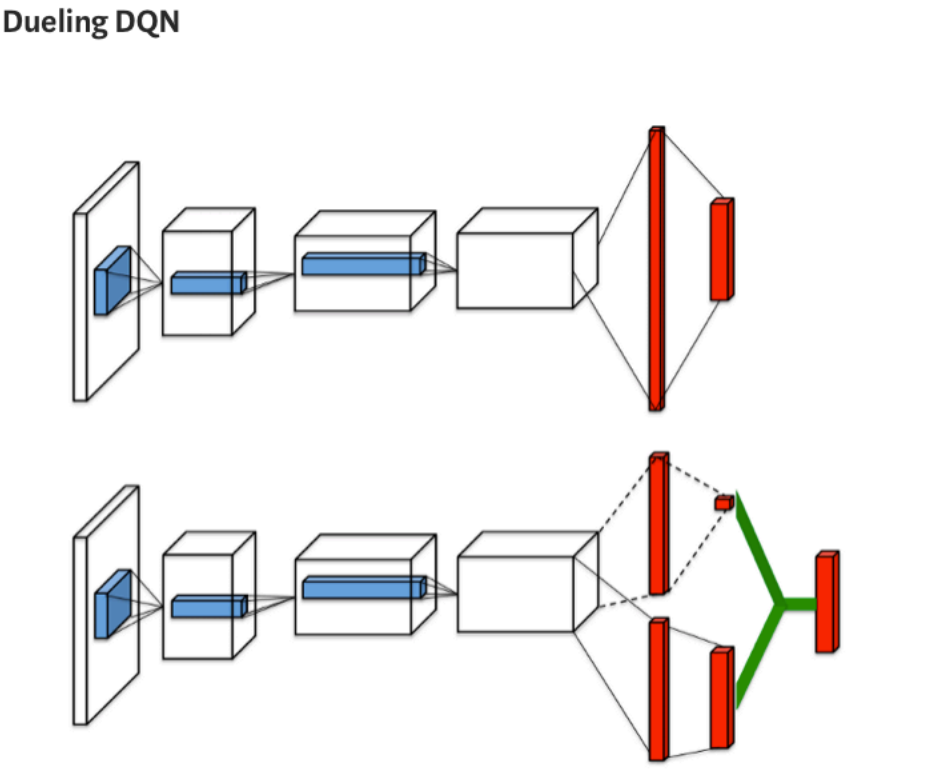
# Learning Algorithm

The learning algorithm used is the Dueling Deep Q Network with the feature having state size as input and 128 layers. Both advantage and value have 128 layers each. Each sequential model has RELU as an activation function (Reference).



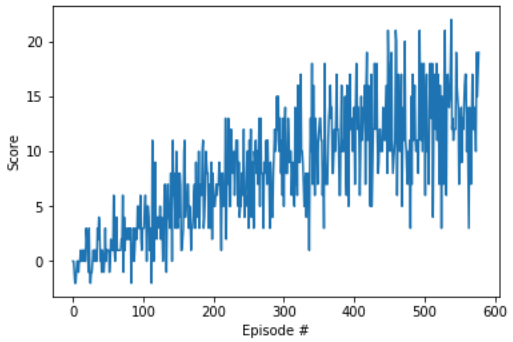

Above: Regular DQN with a single stream for Q-values. Below: Dueling DQN where the value and advantage are calculated separately and then combined only at the final layer into a Q value.

Figure

The code and hypermeters have been adapted from the lesson for DQN. The hypermeters are as follows:
5000: maximum number of training episodes
1000: maximum number of timesteps per episode
1.0: starting value of epsilon for epsilon-greedy action selection
0.01: minimum value of epsilon
0.995: multiplicative factor per episode for decreasing epsilon

# Plot of rewards per episode

```
Episode 100     Average Score: 1.17
Episode 200     Average Score: 4.94
Episode 300     Average Score: 8.17
Episode 400     Average Score: 10.22
Episode 500     Average Score: 12.38
Episode 578     Average Score: 13.05
Environment solved in 478 episodes!     Average Score: 13.05
```




# Ideas for Future Work

Include researching and implementing different improvements to DQN such as prioritized experience replay, noisy networks for exploration, rainbow, quantile regression and hierarchical DQN.