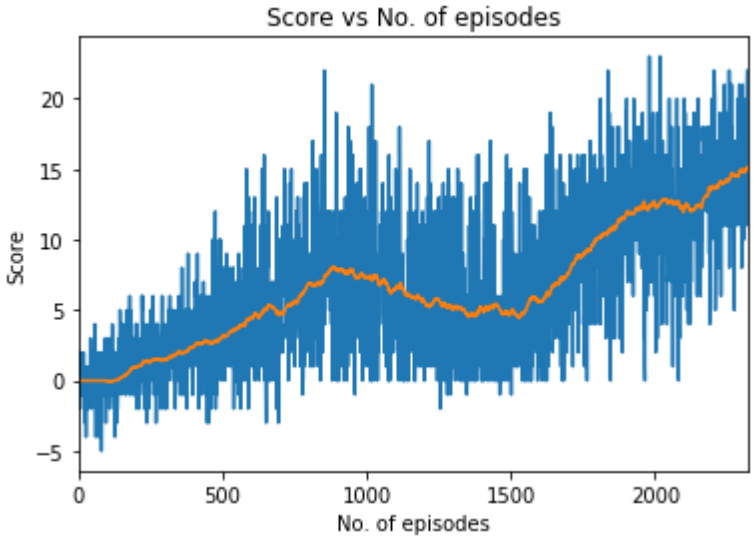
Report

For every 50 steps in a 300 step episode, a step will be picked to train the model. The model takes in 37 inputs and produces 4 outputs, with 2 hidden layers each with 64 nodes using ReLU activation function. The program then runs for 3000 episodes or until an average score of 15 is achieved.



The hyper parameters used are as follows:

BUFFER_SIZE	: 100000
BATCH_SIZE	: 64
GAMMA	: 0.99
TAU	: 0.001
LR	: 0.0005
UPDATE_EVERY	: 50
MAX steps	: 300
Starting epsilon	: 1.0
Ending epsilon	: 0.01
Epsilon decay rate	: 0.995

With the hyper parameters set above, an average score of 15 was achieved in 2311 episodes. At that point, an average score of 15 was achieved by the model. The orange line in the graph is a simple moving average for 100 data points. This is to smooth out the curve and to show the underlying trend of the model.

In the future, different model architecture can be tested to compare the performance of the agent. As for the performance consistency of the model, I plan to change the frequency of the network update to find the optimal frequency with training time in consideration.