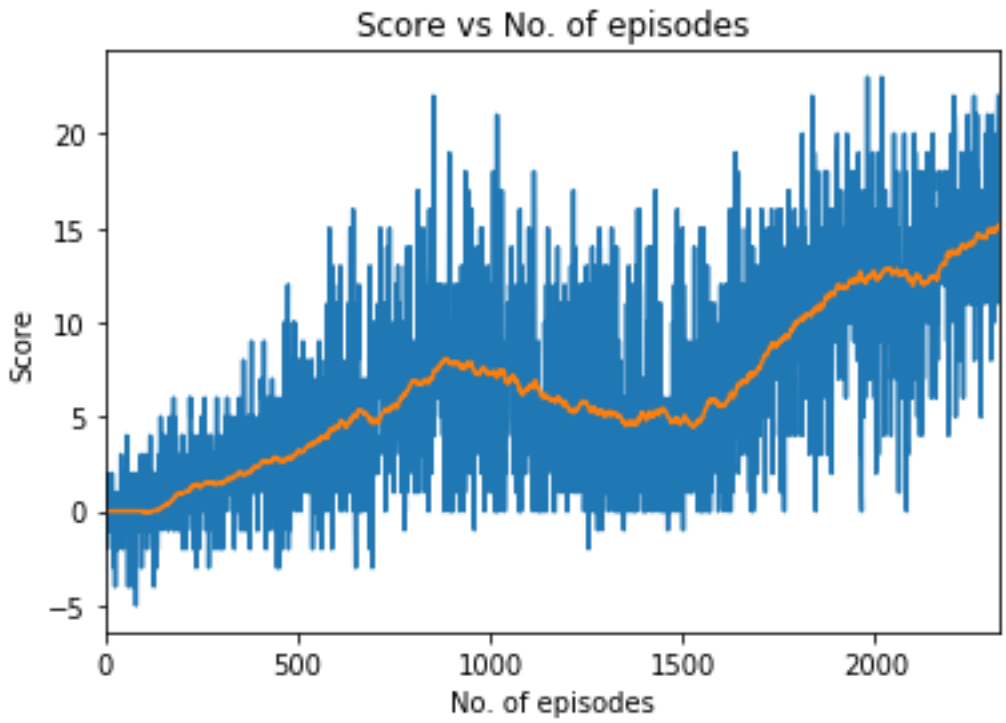
Report

Zhuo Hann Cheah

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. The program then runs for 3000 episodes or until an average score of 15 is achieved.



The graph above shows the performance of the model after 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.