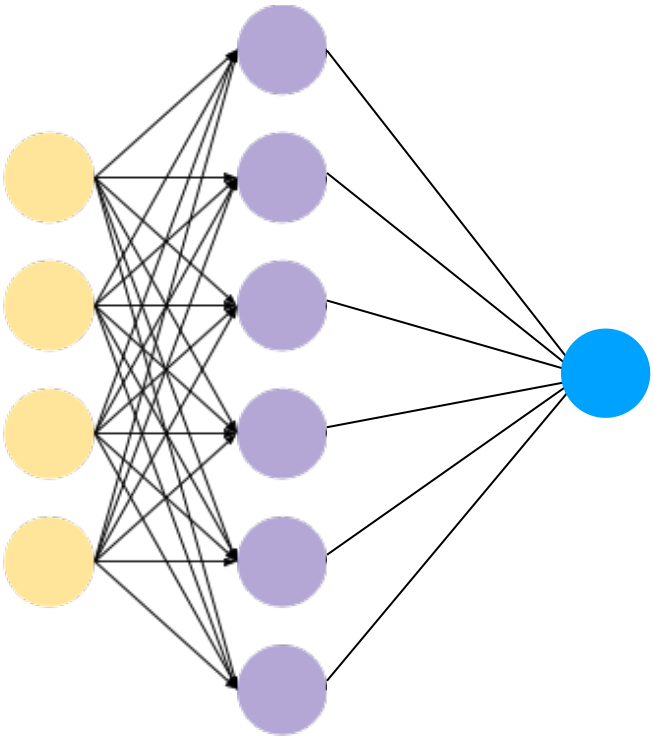
1. Describe your Policy Gradient

在設計 Pong model 時，我一開始選用兩層 Dense layer 來建立 model：



另外參數的設定是如下：

- hidden_layer_size = 200
- learning_rate = 0.0005
- batch_size_episodes = 1
- output_layer_size = 1

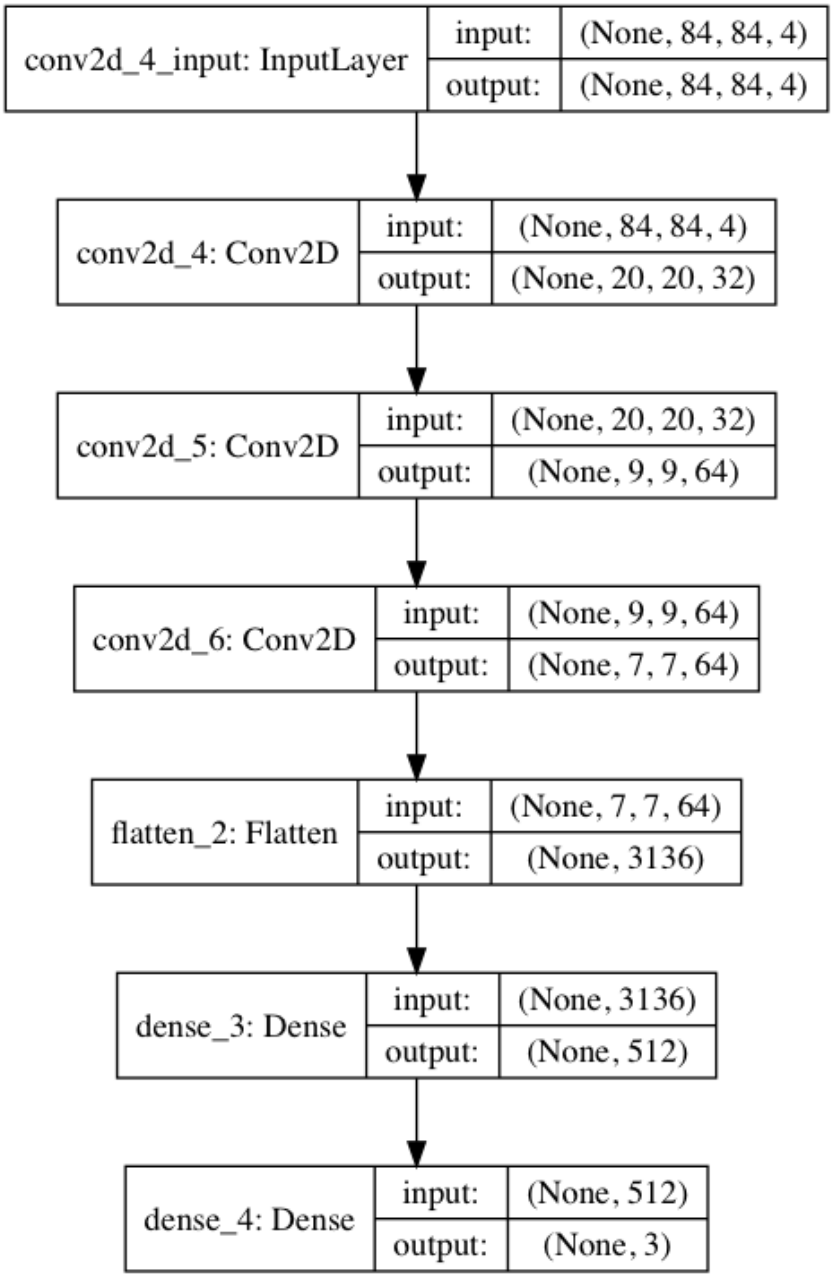
prepossessing 將圖片轉成 80x80 一維的陣列，並且餵進去上述的 model，最後大概 train 了三天左右，總共約 17000 episode，執行 —test_pg 最後平均大概為七分。

另外我有另外嘗試後來助教釋放出來的 model 架構，也一樣 train 了三天左右，總共也 train 了大概 17000 episode 但效果卻沒很好。

2. Describe DQN model

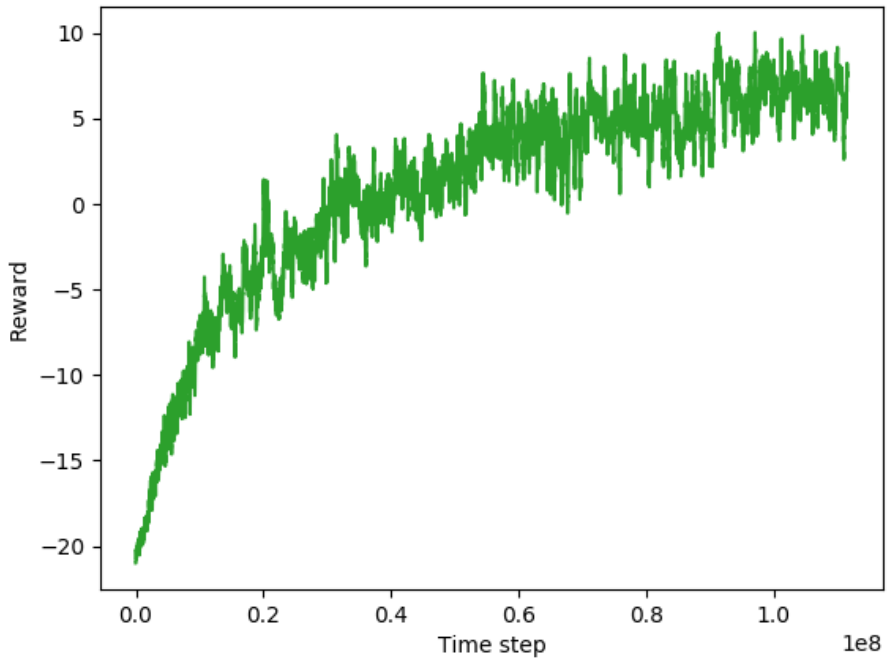
這題的架構參考助教所給，是三層 CNN，一層 Dense，prepossessing 將圖片轉成 84x84x4 的陣列，另外參數設定如下：

- TRAIN_START = 10000
- FINAL_EXPLORATION = 0.05
- TARGET_UPDATE = 1000
- MEMORY_SIZE = 10000
- EXPLORATION = 1000000
- START_EXPLORATION = 1.



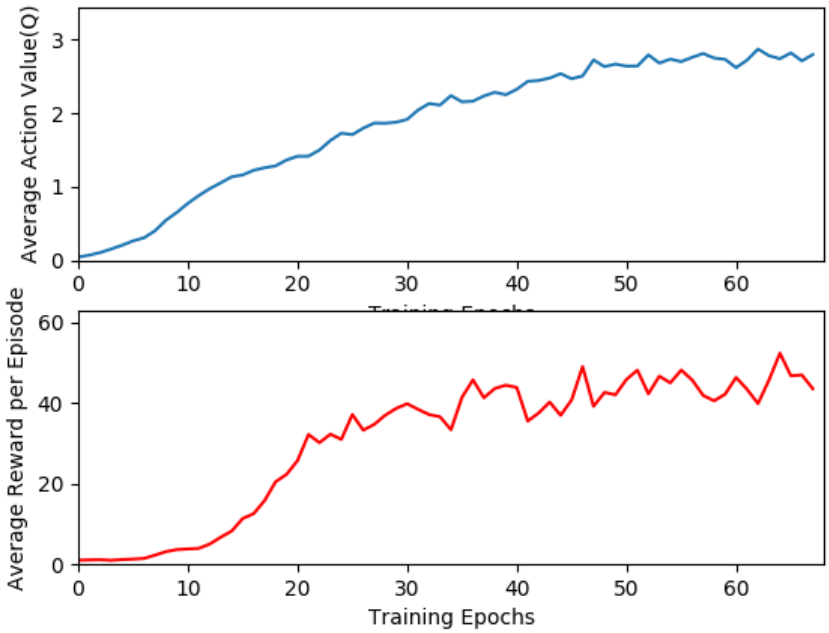
3. Plot the learning curve to show the performance of your Policy Gradient on Pong

兩層 dense model



4. Plot the learning curve to show the performance of your DQN on Breakout

DQN 實作出來後，我大概 train 了整整四天左右，最後卻發現環境設錯，導致 test 的環境無法吻合助教所設環境。以下是我使用 BreakoutDeterministic-v4 所 train 的模型分數。



5. Experimenting with DQN hyperparameters

- Plot all four learning curves in the same graph (2%)
- Explain why you choose this hyperparameter and how it effect the results (2%)

我在 train 的時候有試著把 MEMORY_SIZE 調成二萬，因為我認為能增加 replay 的選擇性，可能最後效果更好，不過當 size 變大時，我訓練就會變得很慢，所以並未完整訓練好。

另外我還有調整 EXPLORATION 參數，一開始我設定了五萬去訓練，結果約莫在 17000 episode 時，平均的 reward 還停留在 0~1 之間。後來採用助教的參數，在 17000 episode 就平均大概有 10 多了。我想探索的步數設多其實效果應該不會比較差，但就是像我只有 CPU 可以用，應該要儘早讓他開始訓練...。