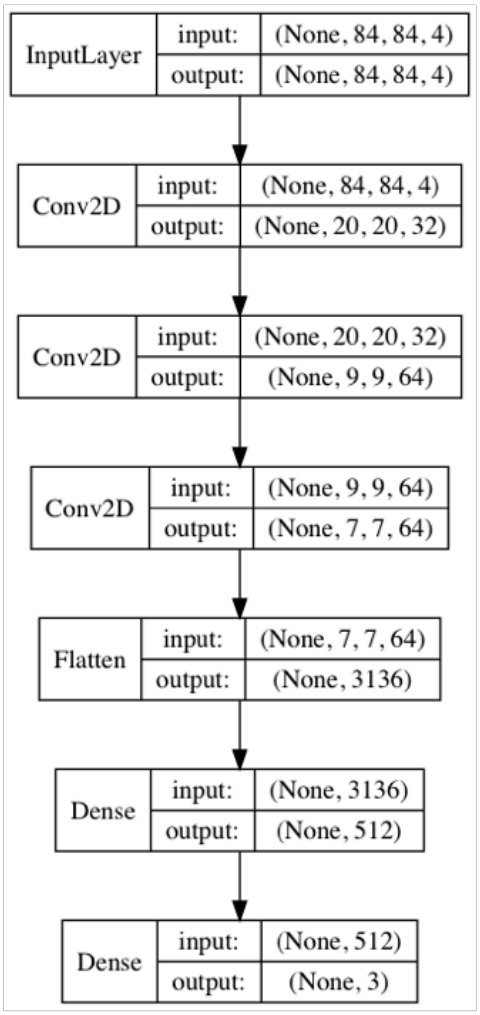
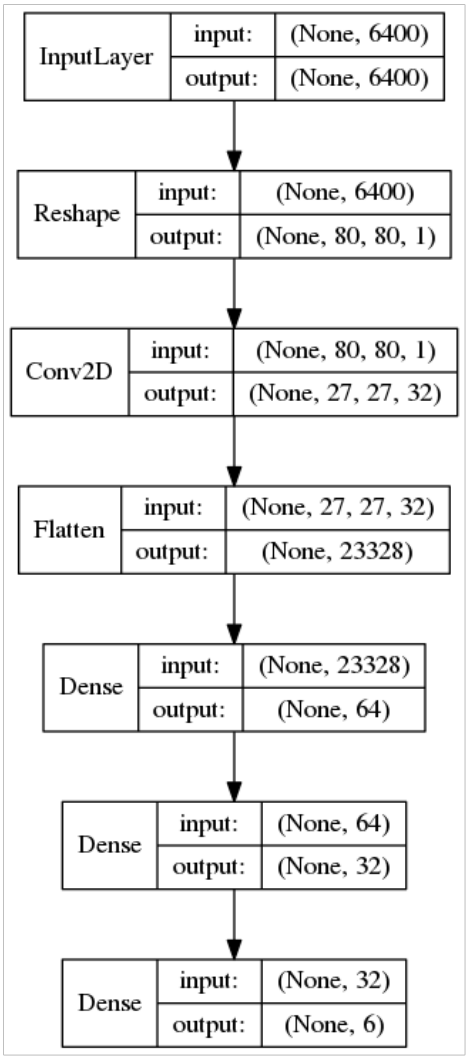
HW3



pong_pg model



breakout_dqn model

兩個model差不多 都是CNN以後接fully connected network

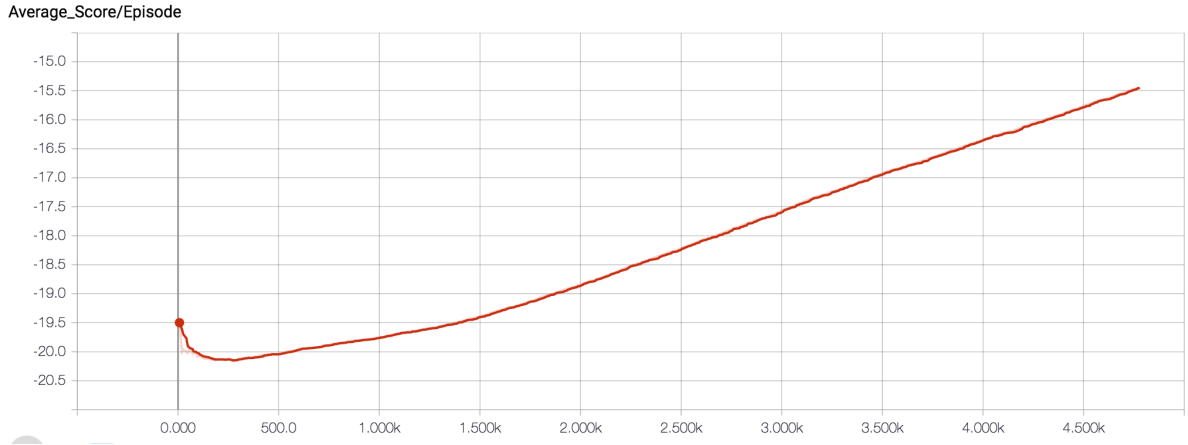
因為pong 有做畫面預處理 所以輸入是80*80

activaton都用relu

dqn model用來預設q value 而pg model用來預測action distribution

所以最後有加上softmax

Learning Curve

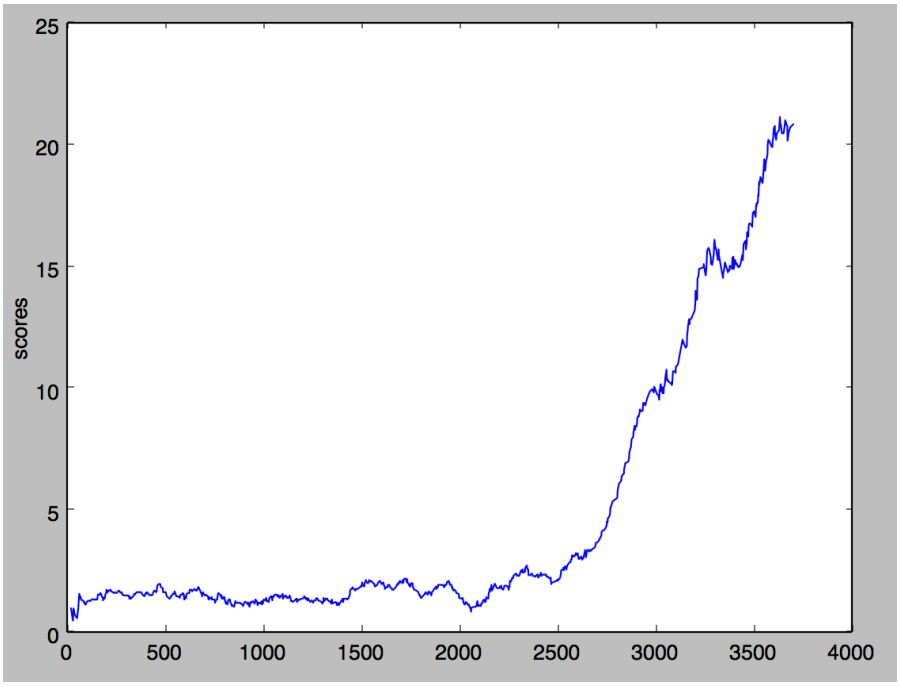


這是上述模型 使用講義上的標準policy gradient做出來的曲線

X軸為episode Y 軸為total reward

不過後來並沒有使用此模型 所以只有節錄一段開始穩定上升的曲線

每個episode只更新一次model (出現21分時完成一個episode)



這是dqn學習曲線

前面會撞牆的很嚴重，總是在0分徘徊，

加入reply memory後才train的起來

不過後面上升得相當快，

另外我在訓練時加入SGD action會練不起來，這是我覺得有點奇怪的地方

而且我的模型沒有出現over estimate q value的情況，在後述與double DQN比較的時候，結果快一點，也許是episode拉長才看得出差異，不過我的機器來不及練，蠻可惜的

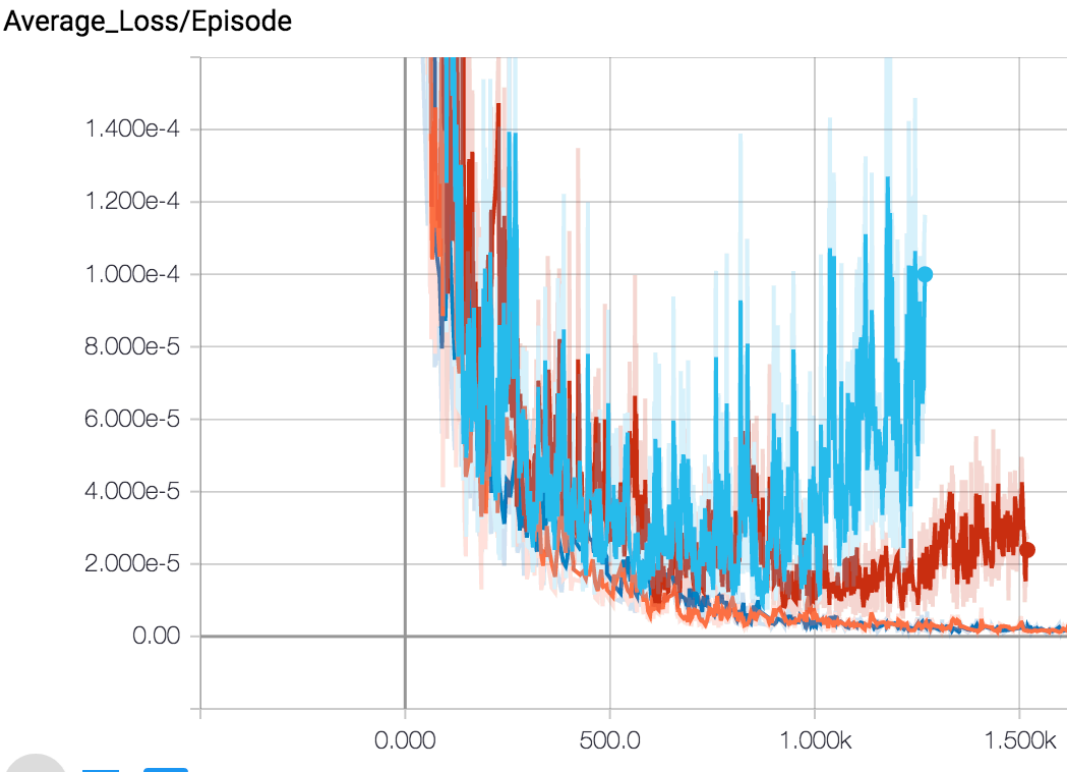
Experimenting with DQN hyperparameters

選用的hyper parameter是target network update frequency

target network update frequency會影響q network的loss

如果更新得太快 會導致q network無法接近target network

所以從前期的loss 震盪就可以看出影響



Y軸為loss X軸為episode Tensor-board smooth:0.5 batch size:32

淺藍色　　:　　1

紅色　　:　　100

橘色　　:　　1000

深藍色　　:　　10000

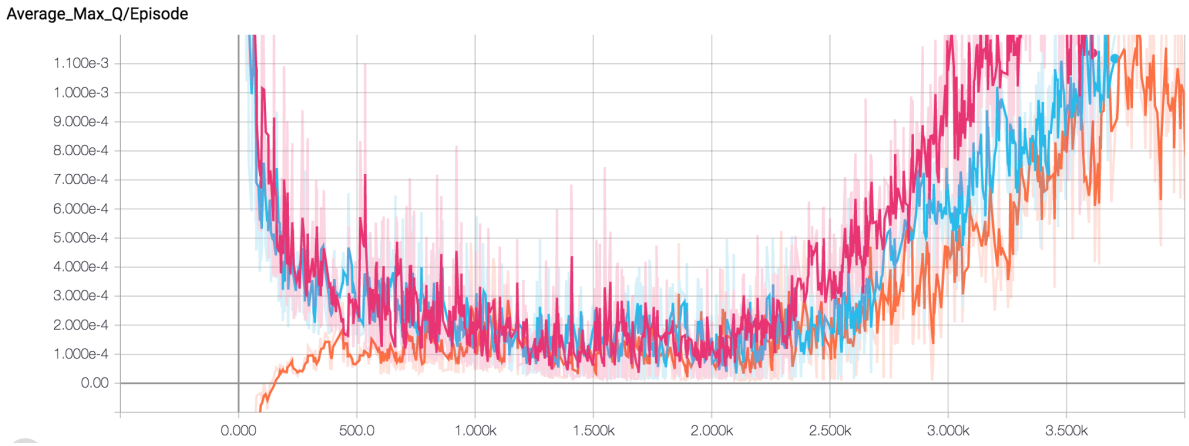
可以看見初期的震蕩由於random initialization所以幅度都滿大 但是在超過1k時,

1000/10000色明顯下降許多，而1/100還在震盪中

Improvements to DQN (2%)

dueling network

double DQN



Y軸為average max q value

X軸為episode

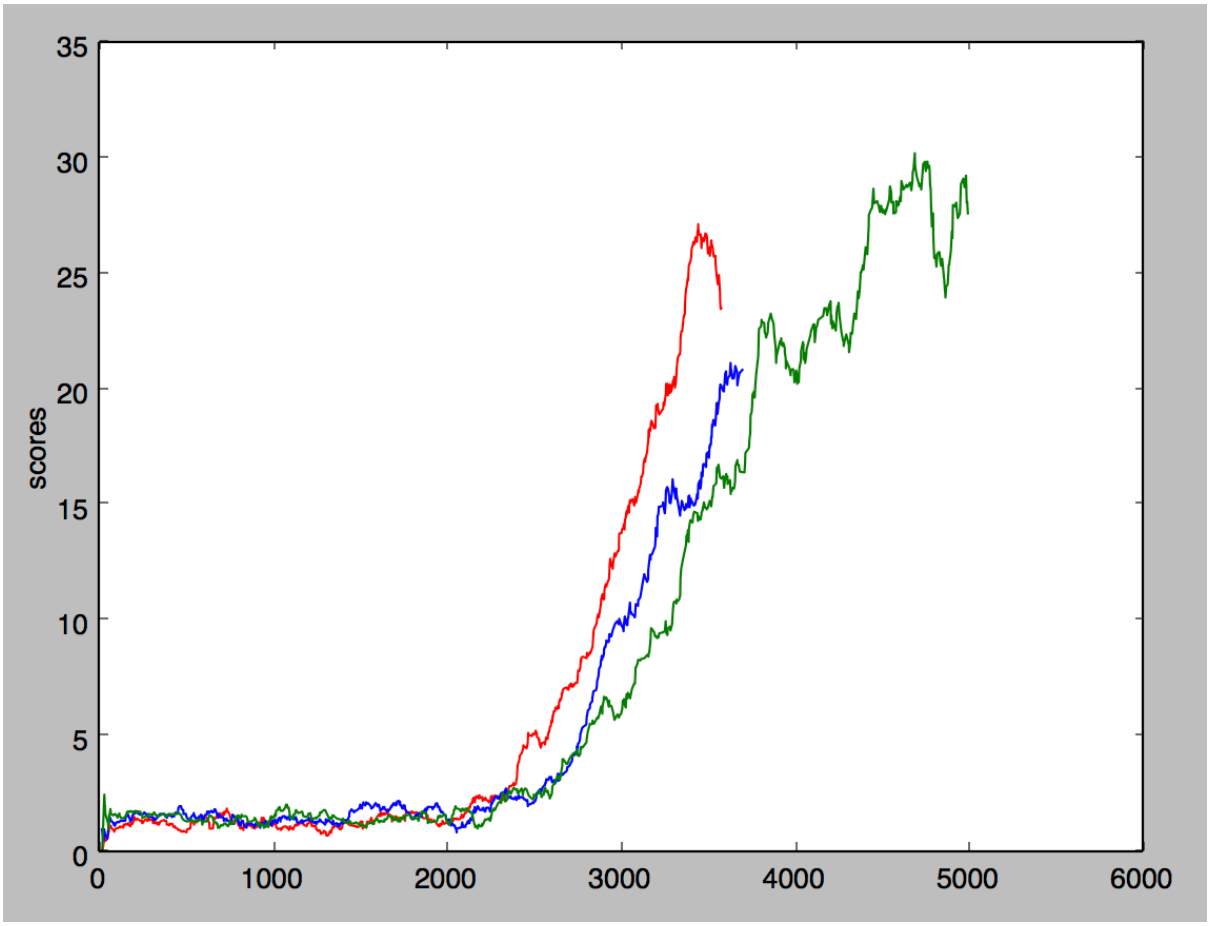
淺藍色為DQN

橘色為Double DQN

桃紅色為Dueling Network

可以看出Double DQN的確Q value較低

但是下頁中看Total Reward比較時 並沒有明顯優勢



X軸為total reward

Y軸為episode

藍色為DQN

紅色為Duel Network

綠色為Double DQN

Double DQN訓練得比較久

可以看出Double DQN的確Q value較低

但是下途中看Total Reward比較時 並沒有明顯優勢