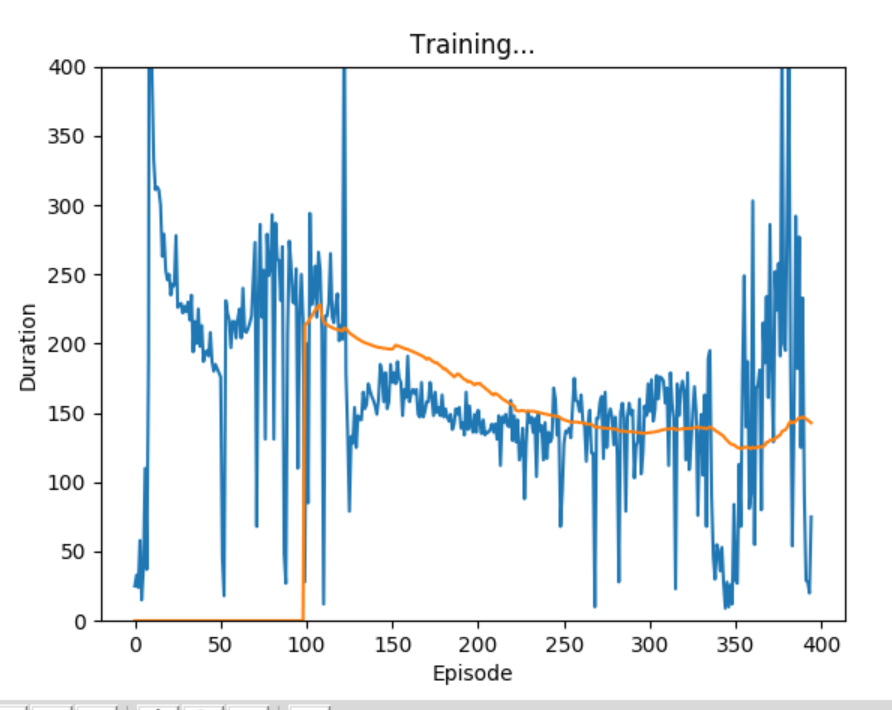
清華大學資工系

大二 學號: A053095 林柏淵

1000+400 episode 前面1000episode沒有截到圖所以再train 400 episode再截下來



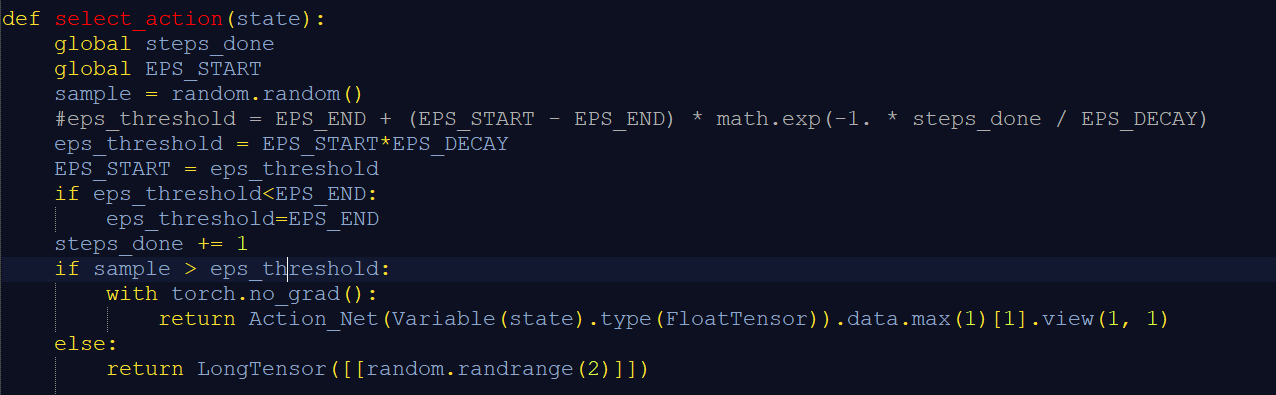
Describe your implement of network structure & Loss function

我用的loss function是adam。然後network structure就跟講義上一樣兩層fully connected，其分別為第一層(4, 32)加上RELU和第二層(32, 2)，其中input size是4由Cart Position, Cart Velocity, Pole Angle, Pole Velocity at Tip組成，再來2個output為向左和向右的指令。

Describe how you implement the training process of deep Q-learning

先select action然後environment再根據action做出next state，再把狀態全部存到memory裡，再來會有兩個network一個負責current q-value一個負責next q-value然後更新其中一個network等到50個episode一個循環再把另一個network做更新。

Describe the way you implement of epsilon-greedy action select method



紅框這邊會判斷說此次的select會交給NN判斷action還是會隨機判斷action的決定。

Describe how the code work (the whole code)

首先會先import openAI寫好的環境gym再來指定要CartPole-v0，再來設定Replaymemory的class用來記錄狀態;

Network本身上面提過的有兩層;

負責畫圖的class用來顯示每個episode總共撐過多久的時間的plot\_durations;

負責選擇要往右還往左的動作的select\_action;

接下來是實際run的部分一開始會先讓environment重置，再來進入一次遊戲回合，迴圈裡面每一個time step都會生成frame而時間流逝後連續的frame就會形成滑車的動畫，接下來就是上面敘述過的training過程(select action等…..)了。