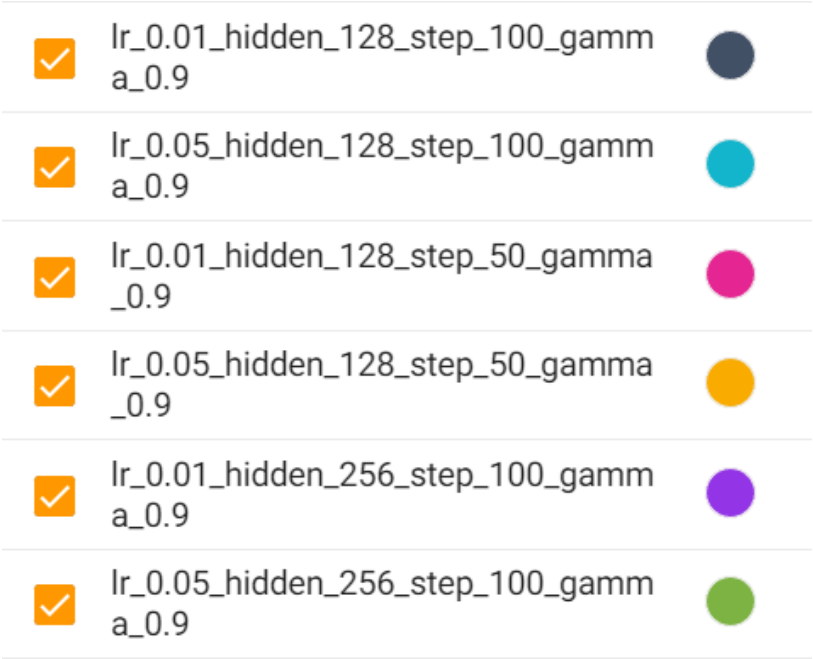
Reinforce Learning HW1

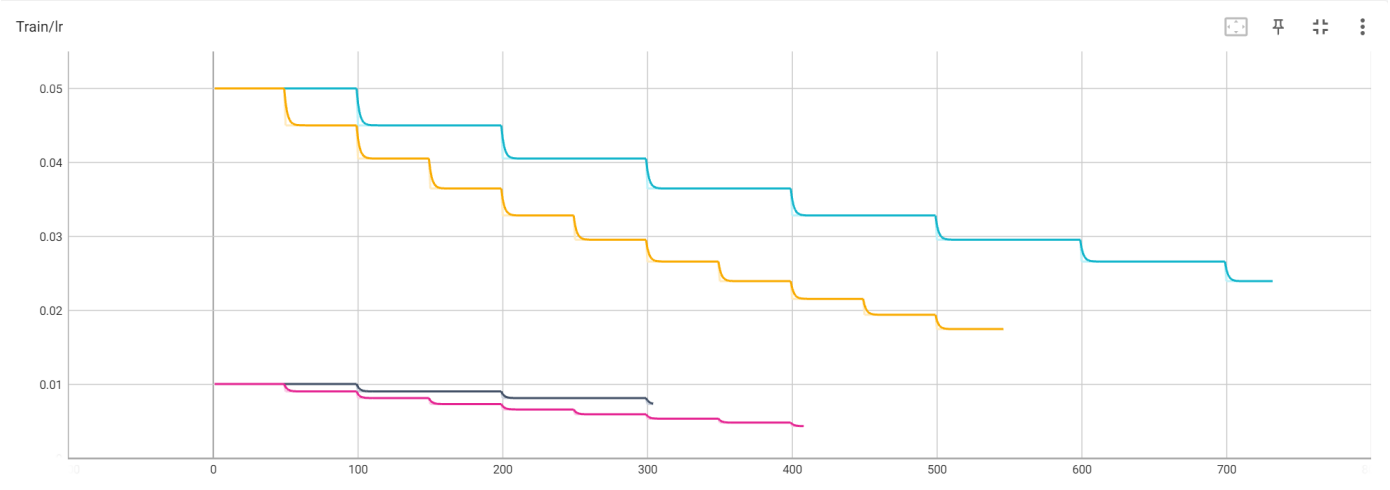
Problem 4

(a.)

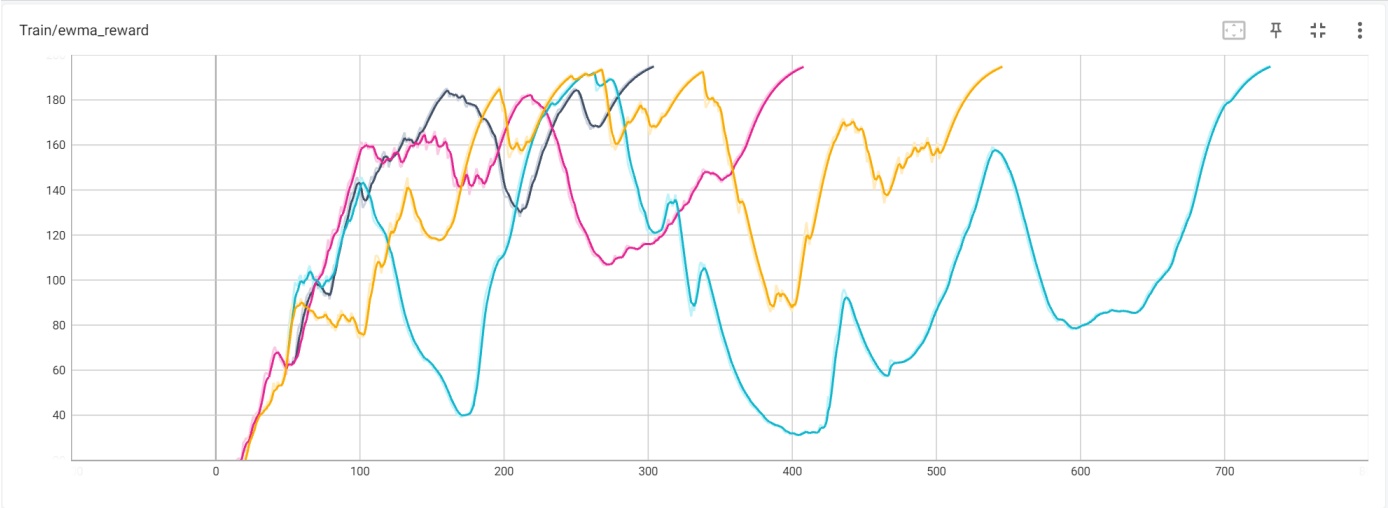
在這個部分我使用了六種不同的參數組合，主要由 Learning rate, hidden layer size 以及 scheduler 的參數的調整，希望可以看出各個參數變化後對於 Learning 的增進又或是趨緩的程度。



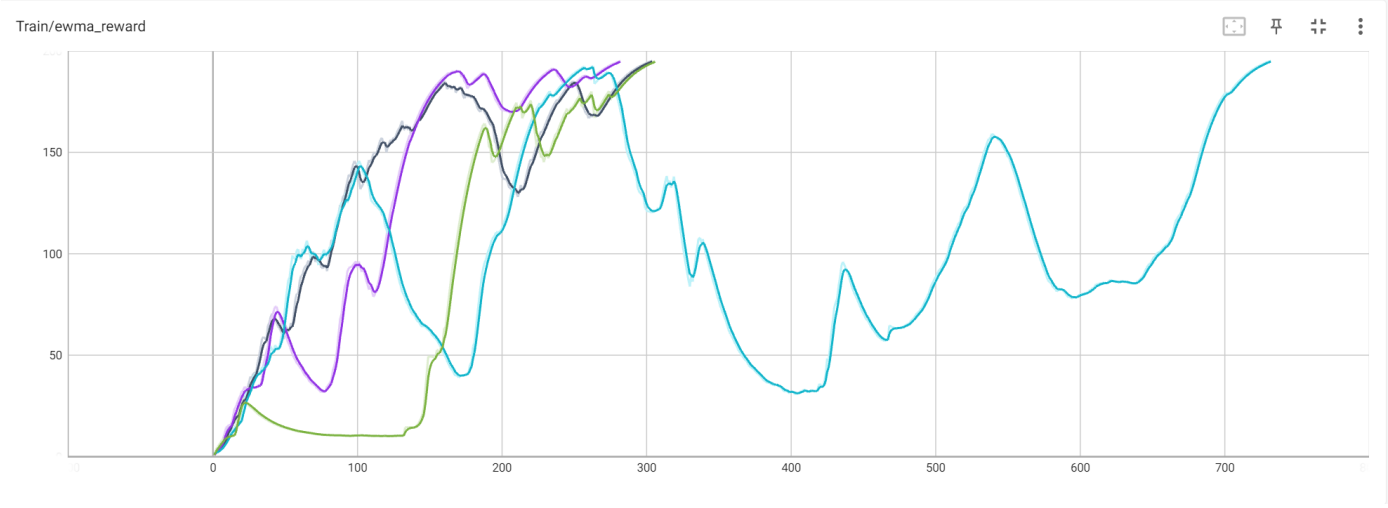
下圖為 不同 Learning rate 隨著 episode 增進的變化



由圖中可以看出在使用更大的 Learning rate 開始進行 Learning 時，會需要更多的 Episode 去完成我們的目標，而在 scheduler 更快的進行 Learning rate 的縮減時，在 Learning rate 較大的案例會有 Learning 進度的增進，但在較小 Learning rate 的案例上，則是會造成副作用，需要花費更多的 Episode 去完成目標。



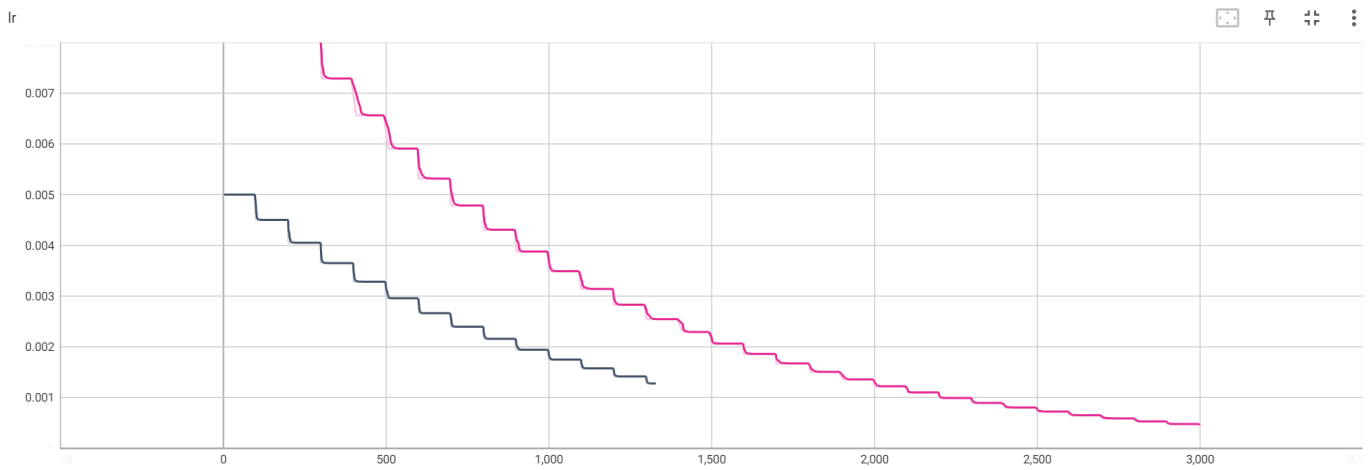
在這張圖所想要比較的則是 Hidden layer 的 size 在不同 Learning rate 上的影響，在這張圖表上可以發現在相同 Learning rate 的情形下，擁有較大 Hidden layer size 的案例，可以較快的達成我們的目標，尤又在 Learning rate 較大的例子上更為顯著。



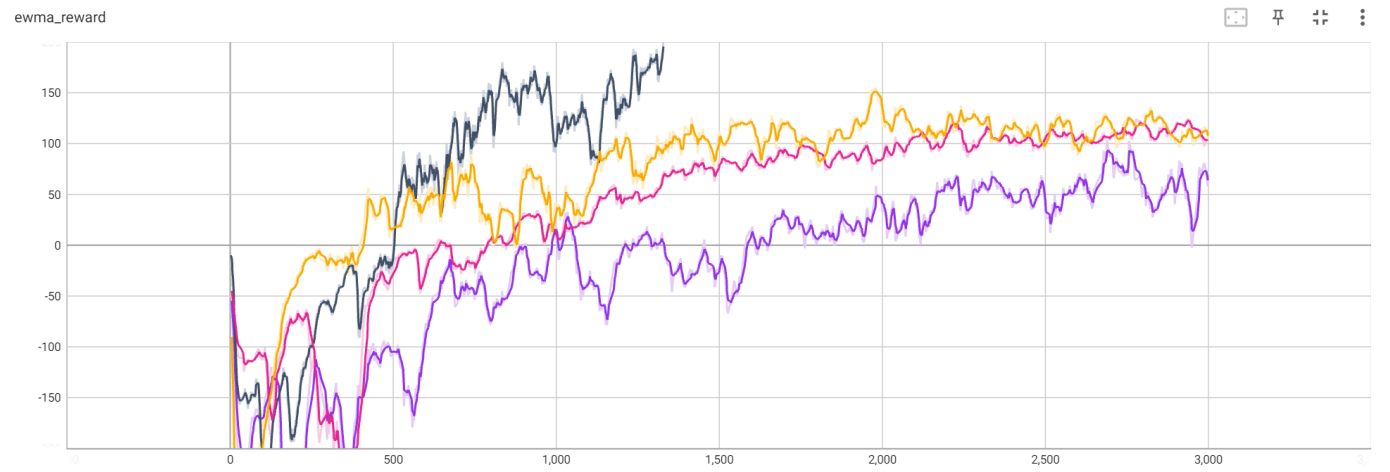
(b.)

在選擇 baseline function 的方面，我主要是用原本就能得到的 State_value 去對每個 State 做 loss 的計算，可以看到其中有一個 案例名字後面有 avg，那個是較為特殊的案例，使用的是所有 State_value 的 mean 值去做為我們的 baseline。

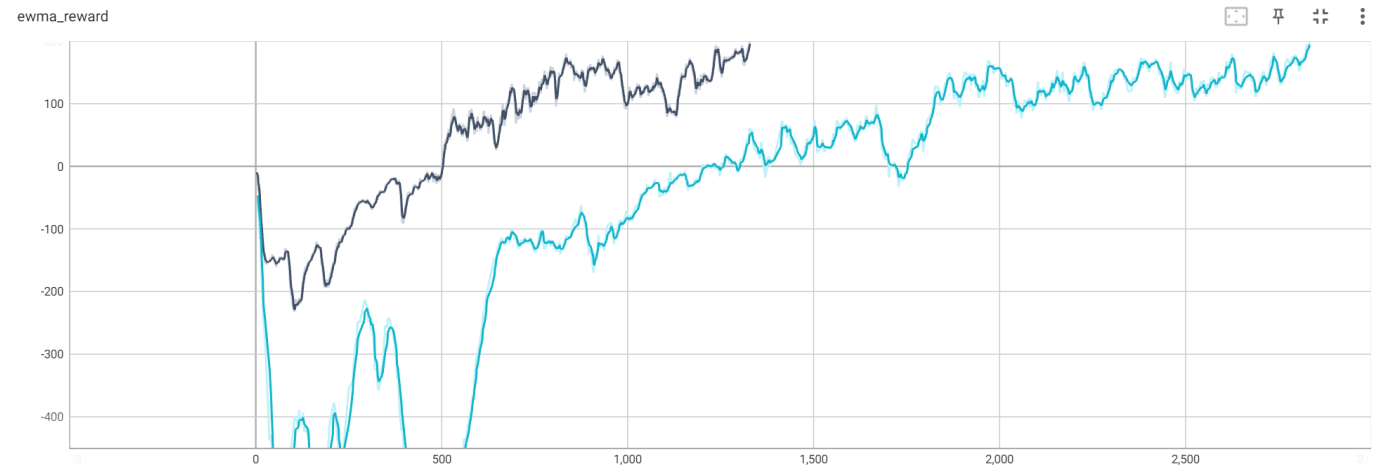
這題所使用的 Learning rate



在這題的實驗中，我只有在設定 Learning rate 為 0.005 以及 hidden layer size 為 256 時才成功的達成目標，使用其他的幾種參數去進行 Learning 時，通常 average reward 到 100 多就開始跳動了。與此同時，好像 hidden layer size 為 128 的也不能真正很好的去映射這個問題，因此在 Learning rate 相同的情形下，最後也沒有如同 hidden layer size 256 的那般達成目標。

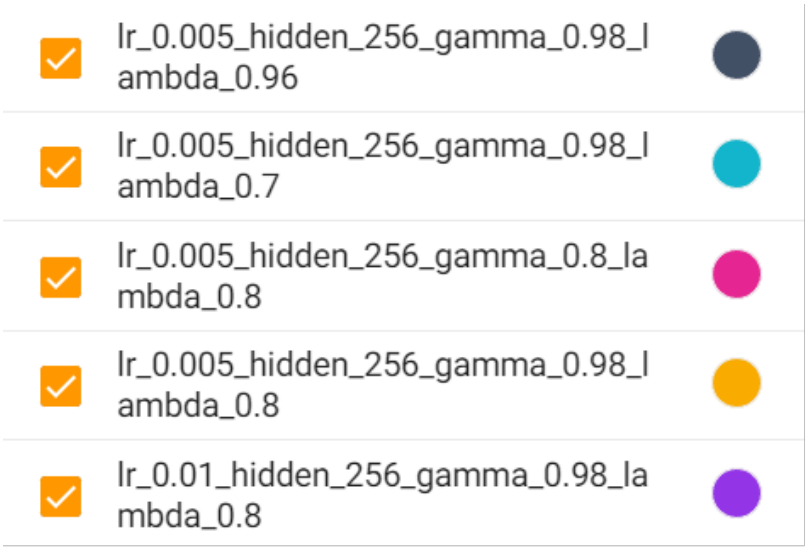


這個圖表是用來展示在使用兩個不同的 baseline function 時，兩組相同參數的模型的收斂速度。即使是利用每個 State_value 的 average 去做為我們的 baseline，在經過較長時間的訓練後，還是可以有效的達到我們的目標。

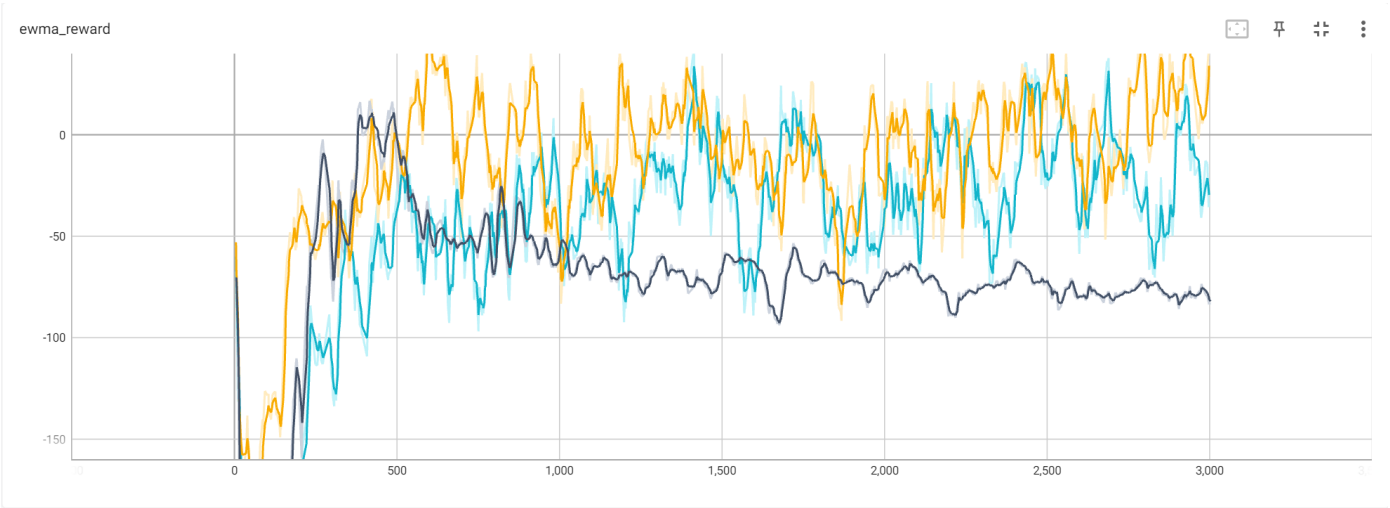


(c.)

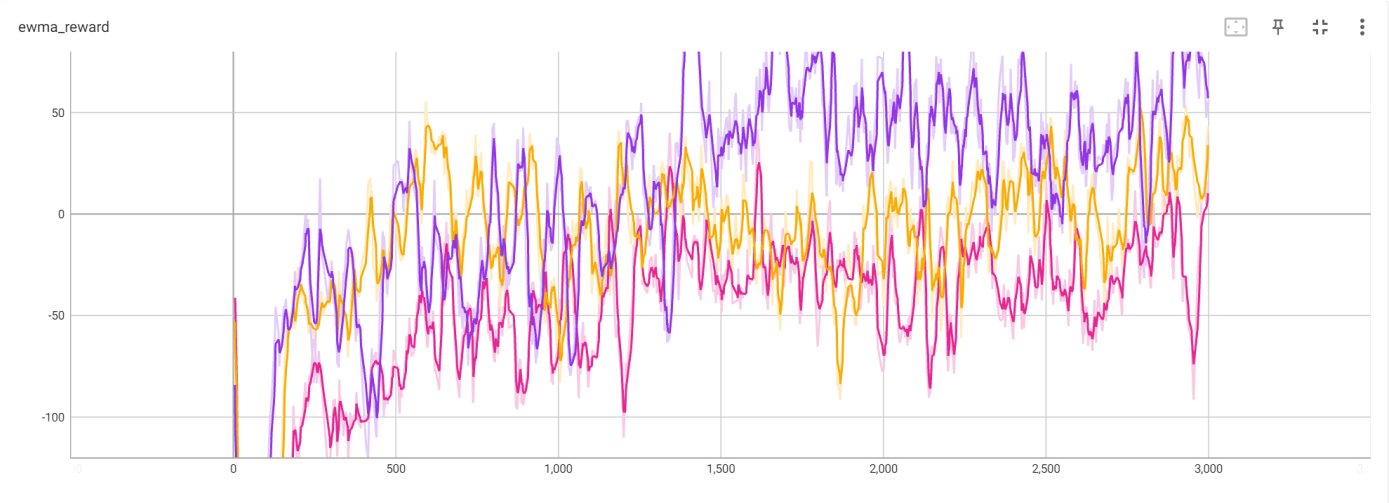
在使用 GAE 的這個小題中，我沒有成功的訓練出可以完成 LunarLander 的模型，我嘗試的參數大多數都上升到 0 附近後便開始來回波動，沒有再繼續往上的趨勢。因為在 (b.) 的成功經驗，因此在本題中，我多數選擇使用在 (b.) 已經成功過的參數，並以此為本來嘗試改變我們 GAE 中的 λ 與 γ 值。



在這個案例中，我嘗試比較改動 GAE 中的 λ 時，對於 Reward 的影響，主要嘗試了在原 paper 中，對於 CartPole 有較好表現的 0.96 以及之後隨機嘗試的 0.8 與 0.7，從圖中可以看到 0.8 與 0.7 的結果相較於 0.96 成功不少，但也只是收斂在 0 附近，沒有持續上升的趨勢。



而下面這張圖，比較的則是不同 learning rate 與 γ 對於 Reward 的影響，其實可以發現好像 Learning rate 為 0.01 時，相較於 0.005 更為成功一點，而 γ 的往下改動則看起來像是帶來負面的效益，沒有真正的幫助到訓練的結果。


ewma_reward
50
0
-50
-100
0
500
1,000
1,500
2,000
2,500
3,000