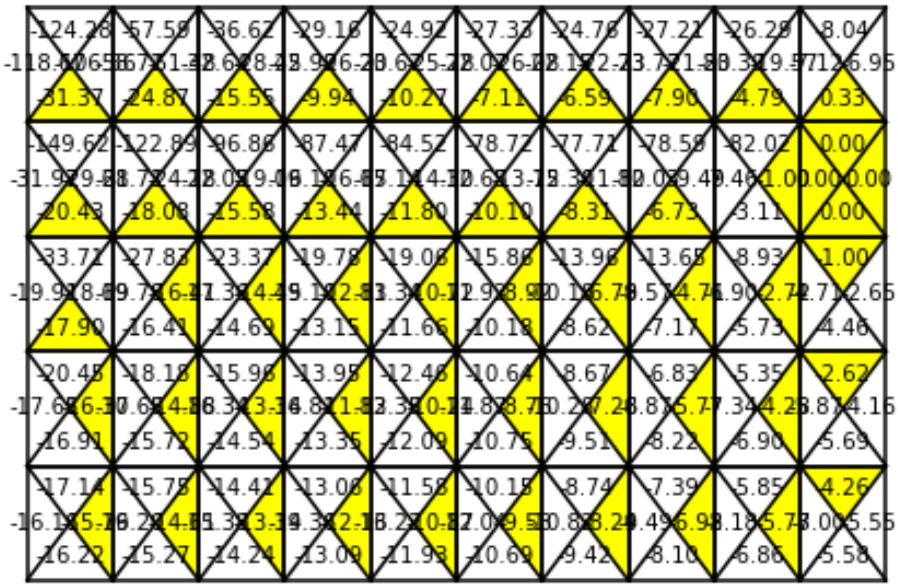
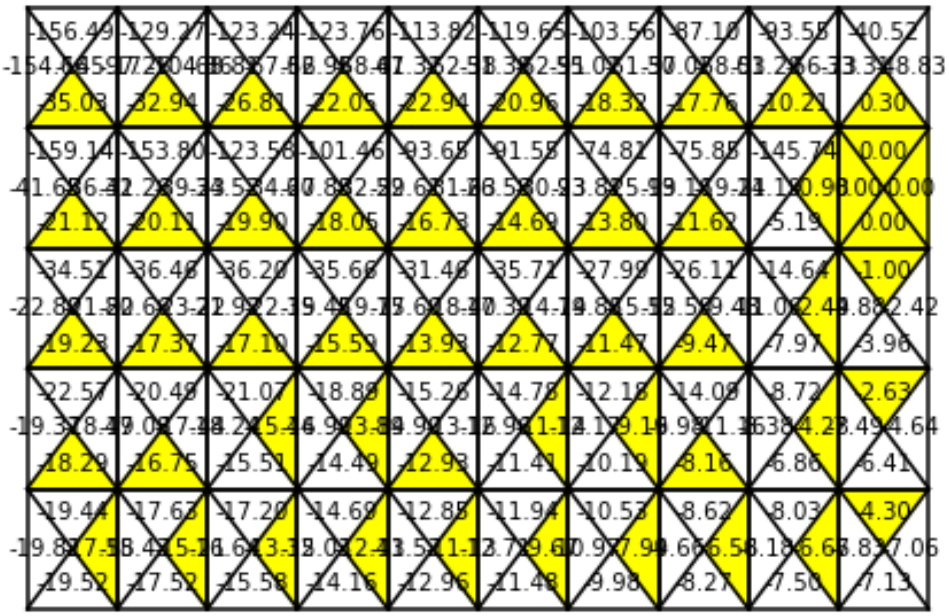
# Experiments and Analysis (40%)

1. Plot the Q-values of Sarsa and 5-steps Sarsa, and explain your result. (15%)

**Sarsa**



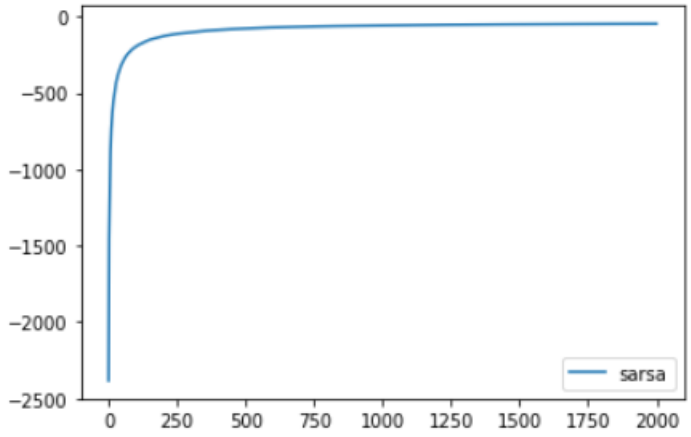
**5-steps Sarsa**



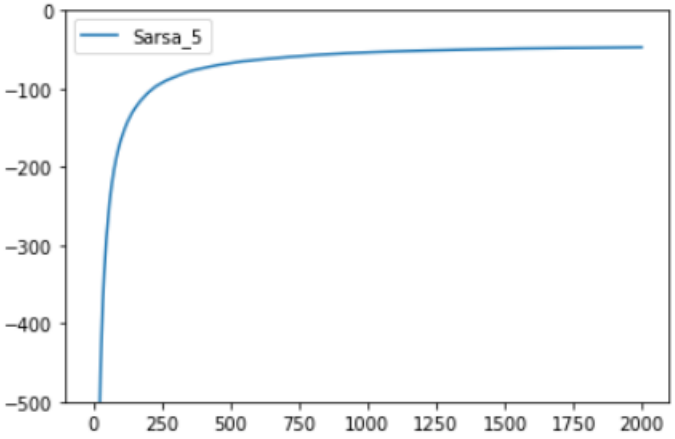
因為同時取得 5 步的資訊，所以得到有用的資訊的機率比較高，n 步 Sarsa 比 Sarsa 容易選擇更保守、更安全的路徑，因此距離 Swamp 越遠越安全，而 on-policy 來說可能最遠距離就是最佳路徑。

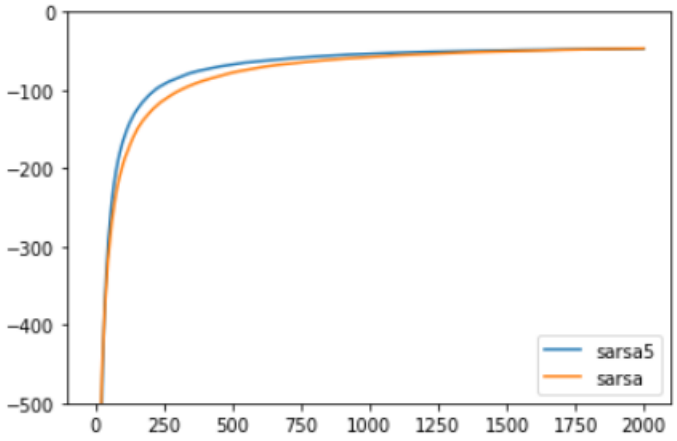
2. Plot the average returns of Sarsa and 5-steps Sarsa, and explain your result (15%)

**Sarsa**



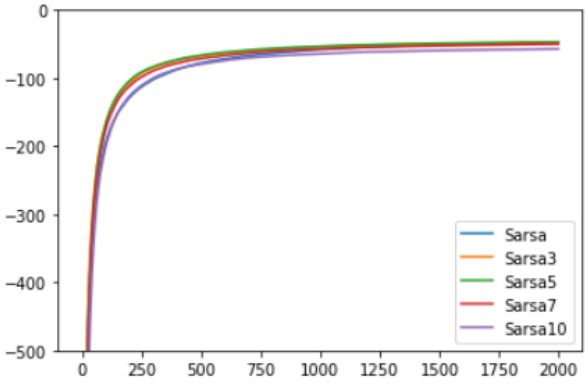

**5-steps Sarsa**

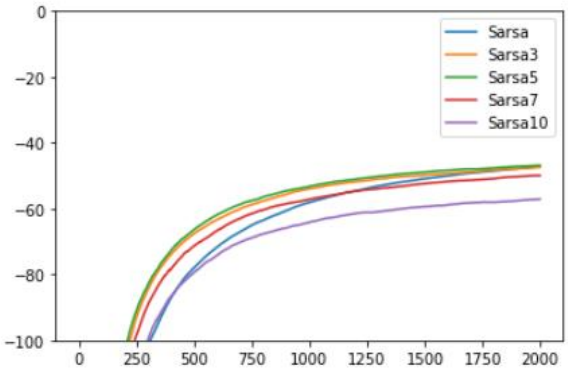





從對比圖中可以看出，5 步 Sarsa 比 Sarsa 學習速度快，因為 5 步 Sarsa 累積了 5 步的資訊才更新價值，只要 5 步裡包含有用的資訊，此狀況機率較高，代理人每一步就能進行更新，相較之下，Sarsa 的每一步都必須包含有用的資訊，此狀況機率較低，代理人才能每一步進行更新學習。但當學習時間拉長後， Sarsa 逼近 5 步 Sarsa 的平均報酬。

3. Varying n-steps and get average returns, then compare by overlap the plot (10%)



從第二張放大圖中可以看到，學習速度來看的話 Sarsa5> Sarsa3> Sarsa7> Sarsa> Sarsa10，一般來說 n 步 Sarsa 的步數越多，學習速度越快，但是如果 n 過大趨近無限大會接近於蒙地卡羅演算法，其效能和學習行為都會差不多，因為累積過多的資訊才做更新，因此造成學習速度和平均報酬都下降，因此 n 的選擇也非常重要不能太大也不要小，才能達到最大效果。