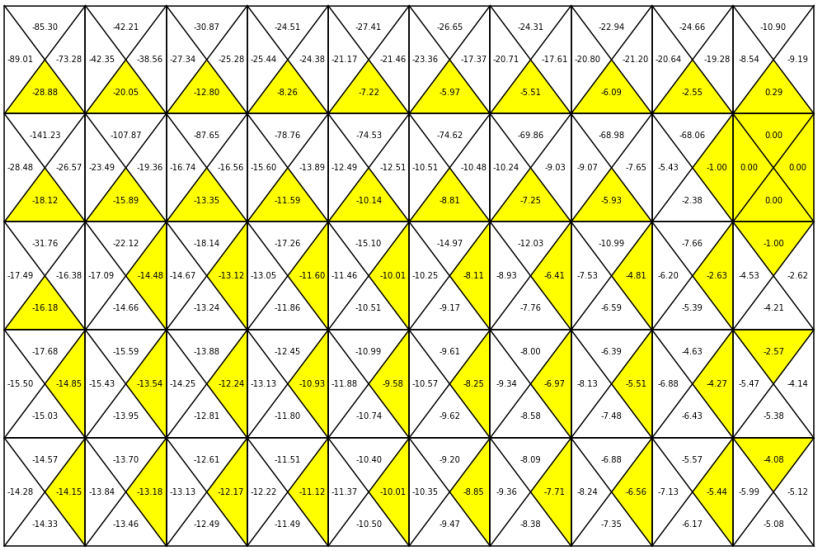
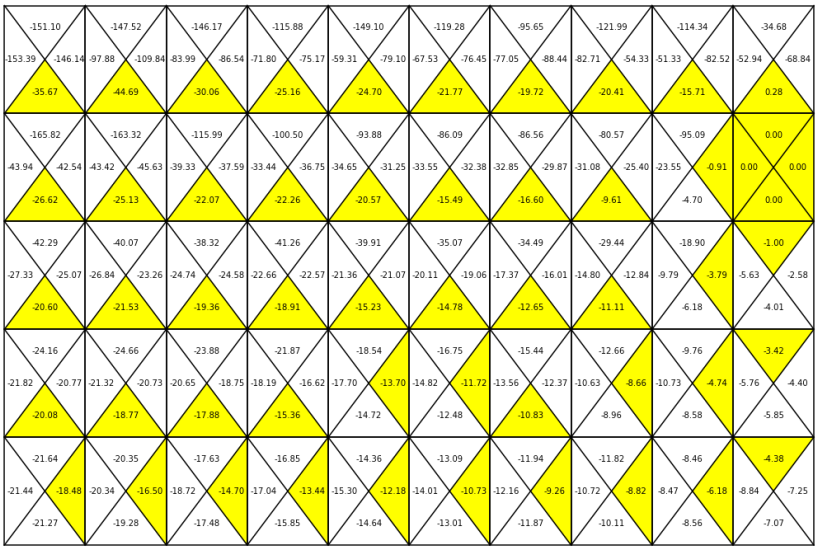
Experiments and Analysis(40%)

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

   Sarsa：

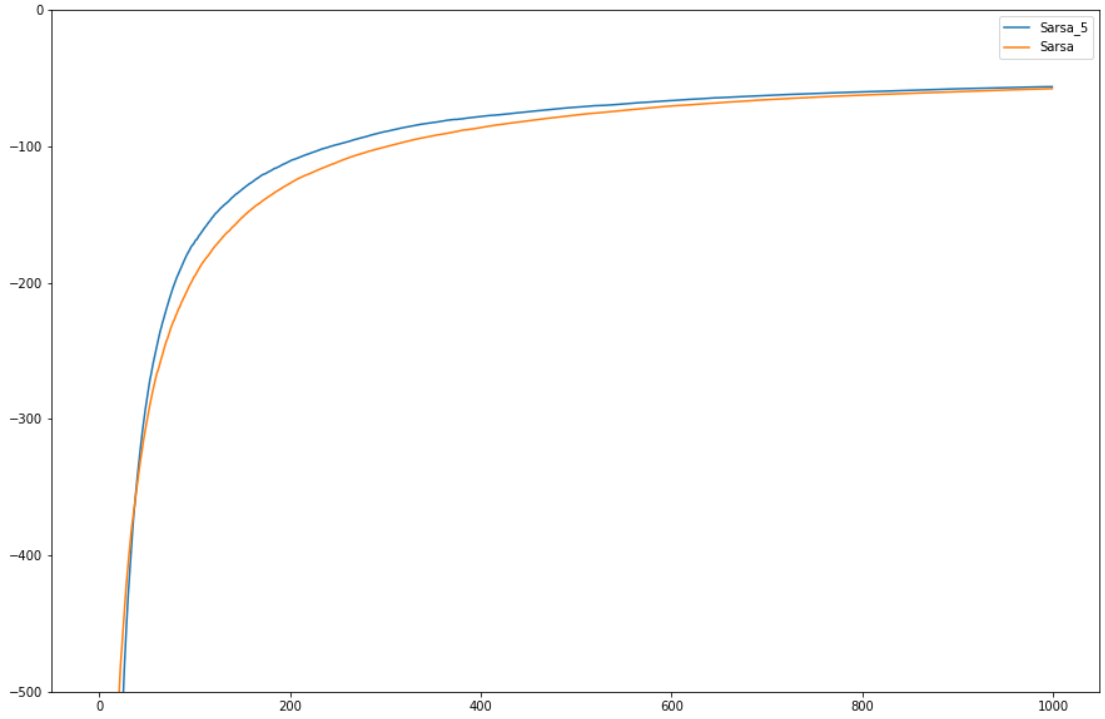


   5-steps Sarsa：



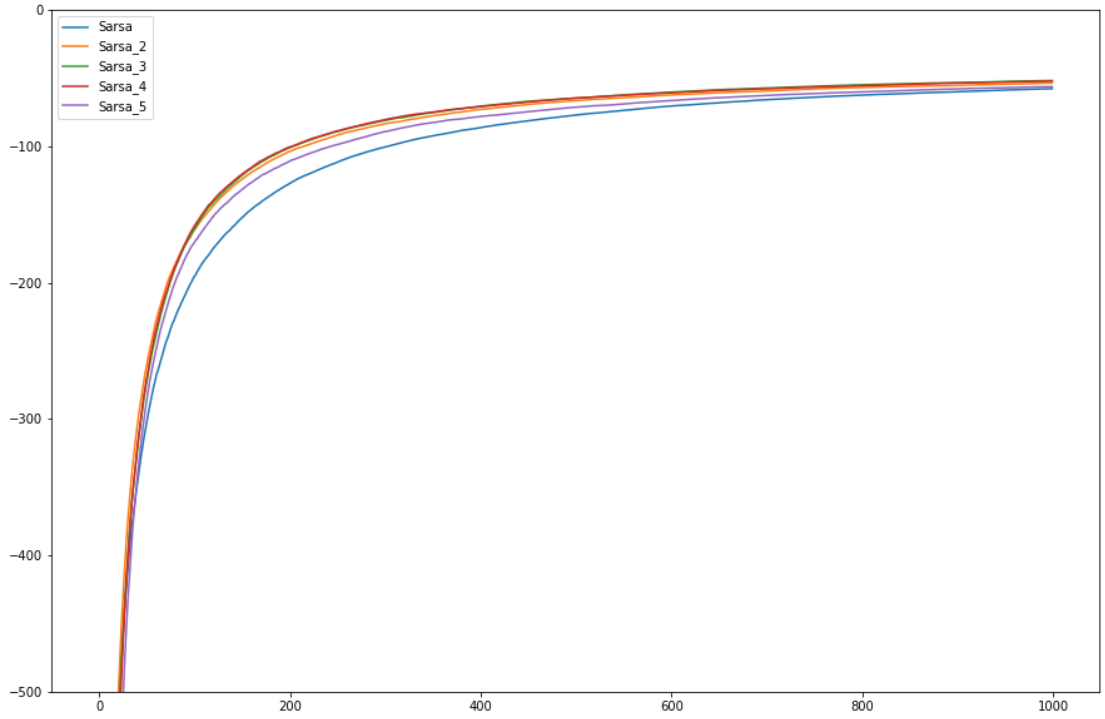
   5-steps Sarsa 同時會取五步的資訊做學習，取得有用資訊的機會會比 Sarsa 高，且 5-steps Sarsa 和 Sarsa 是 on-policy，所以 5-steps Sarsa 會選擇走得比 Sarsa 更遠離 Swamp，也就是更安全的路徑。

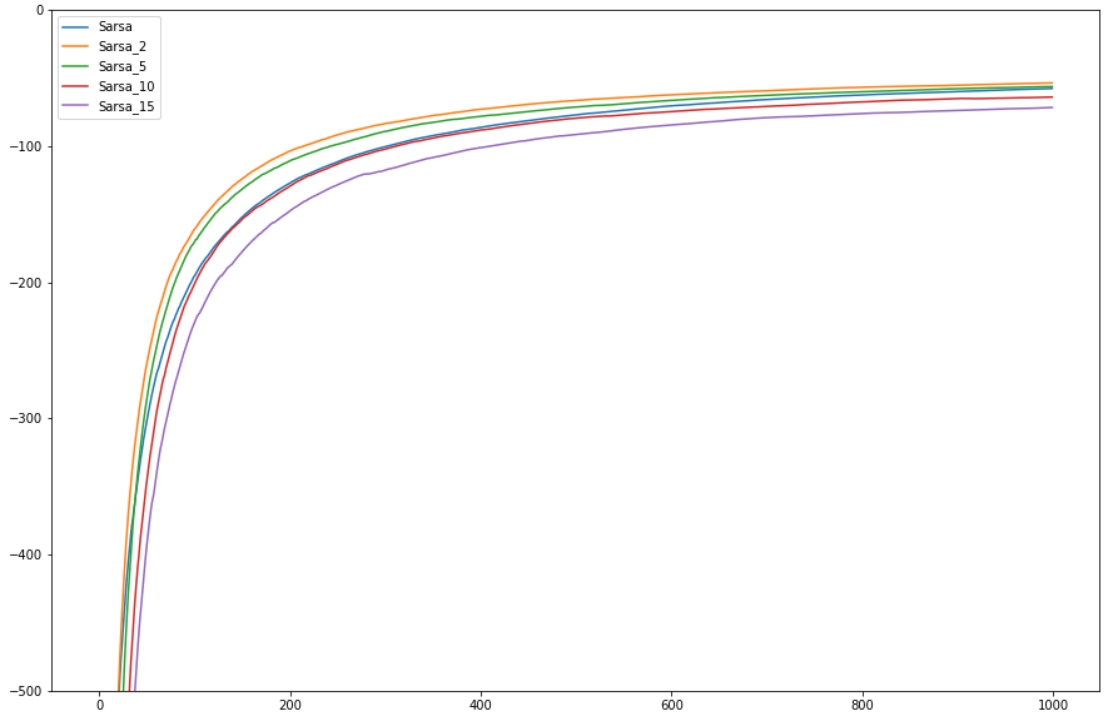
2. Plot the average returns of Sarsa and 5-steps Sarsa, and explain your result(15%)



Sarsa 在最後一個狀態動作配對才有動作價值改變，而 5-steps Sarsa 在最後五個狀態動作配對皆有動作價值改變，也就是說 5-steps Sarsa 一次會學習五步的資訊，所以 5-steps Sarsa 的學習速度較快，但學習時間拉長後，Sarsa 的平均報酬也是會逼近 5-steps Sarsa。

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





一般來說，n-steps 的步數越多，學習的速度會越快，也就是收斂的速度會越快，但步數過多也可能導致相似於蒙地卡羅，已經累積太多資訊才更新，就會有造成平均報酬下降的狀況發生，例如圖中的 15-steps Sarsa。