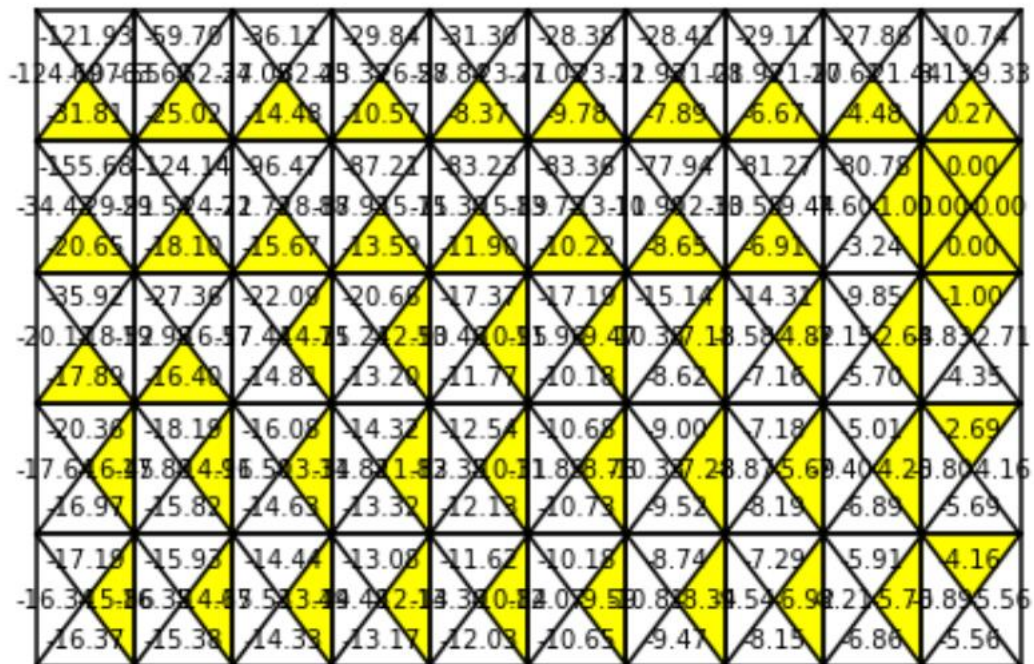
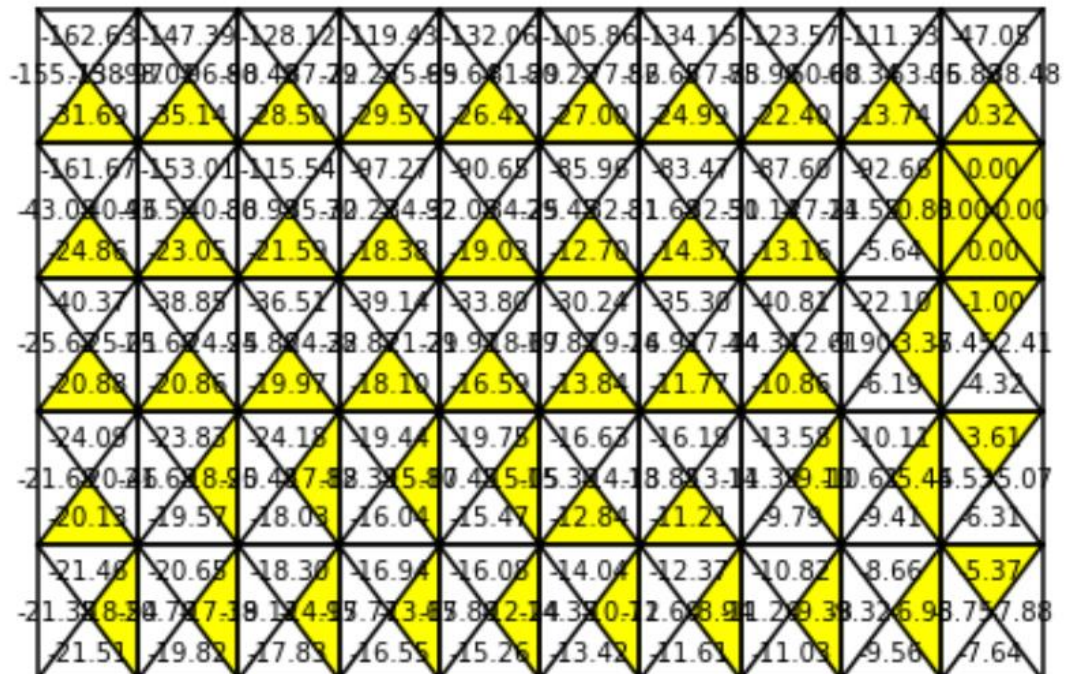


1. Plot the Q-values of sarsa and 5-steps Sarsa and explain your result

Sarsa Q-values



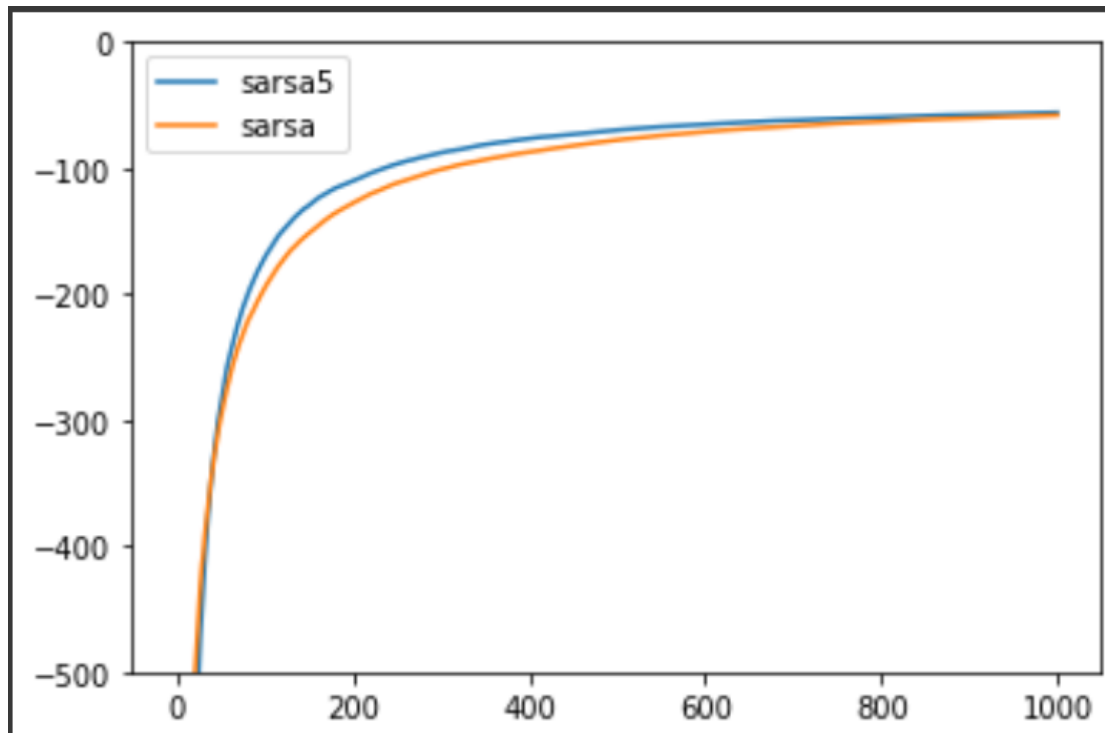
5-steps Sarsa Q-values



在相同的 epsilon 下，5 steps 走到 swamp 的機率會比 sarsa 來的大，因為必須等到五步之後才做更新，所以她很有可能會走到 Swamp 中，所以他會走較保守的路線，而 sarsa 則為做一步更新一步，所以它有較低的機率走

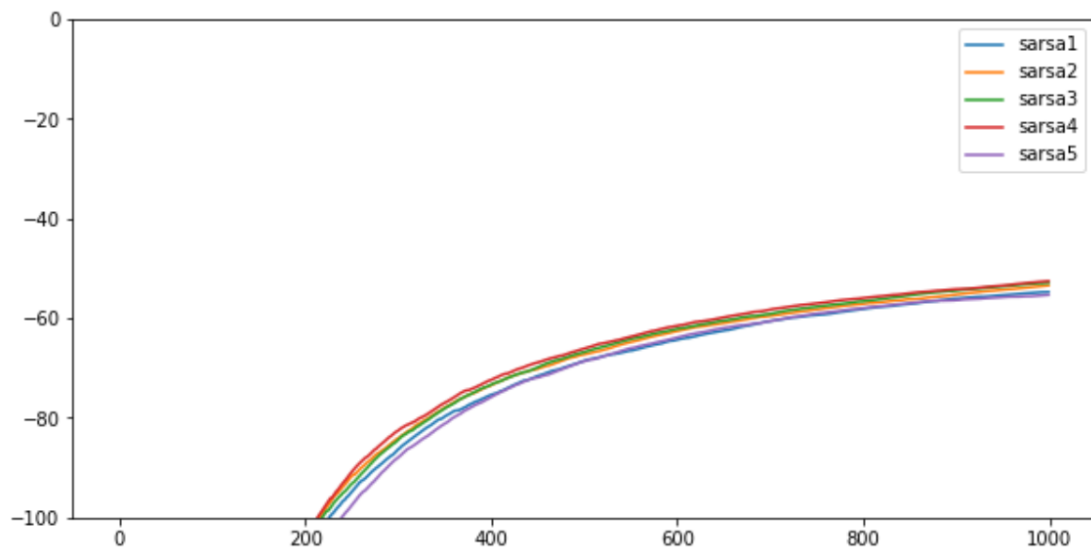
到 Swamp 中，所以會選擇較快一點的路徑。

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



收斂速度上，Sarsa5 是比 Sarsa 來的快，但是隨著時間增加(步伐增加)，則兩種方法會趨近於差不多的值

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



Sarsa3 和 sarsa4 通常都會是最好的兩個，最差都會是 sarsa5 和 sarsa，所以在這個環境中，3 或 4 steps 是可以得到最好的平均 reward，應該是跟環境有關。