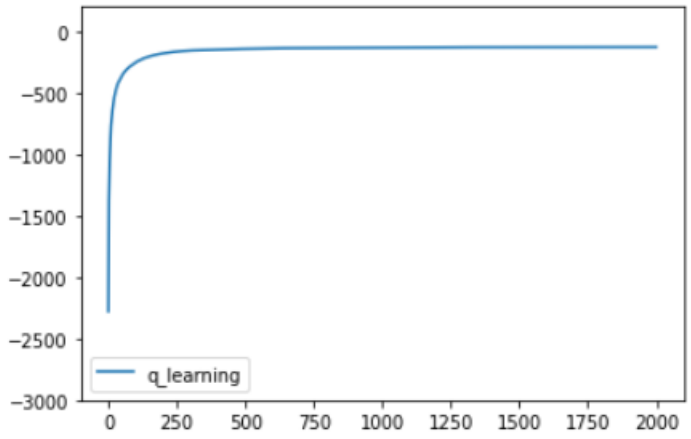
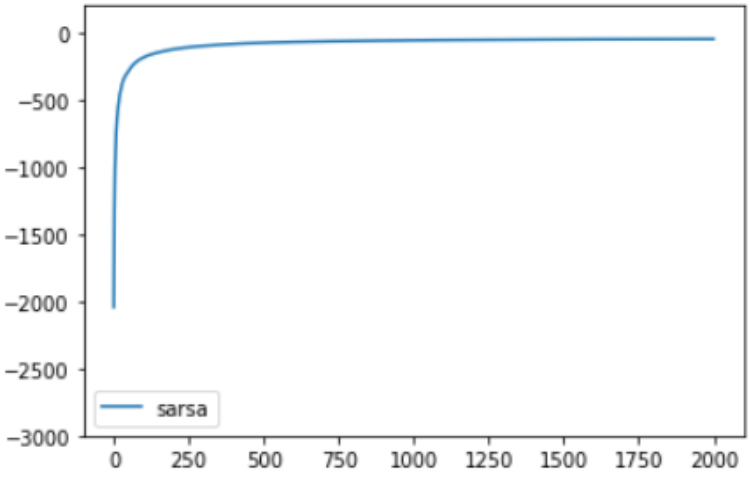
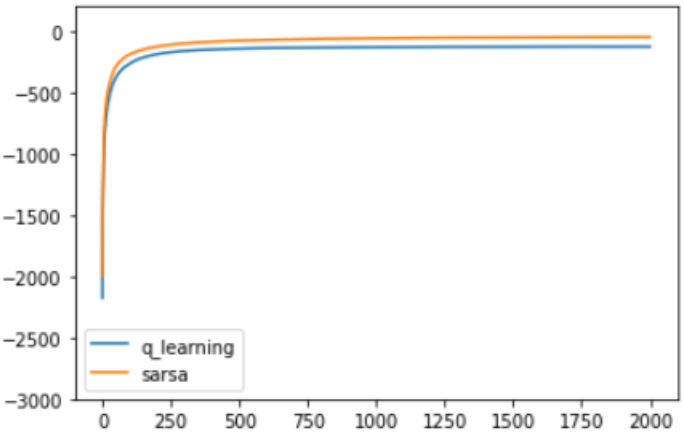
Experiments and Analysis(40%)

1. Plot the average rewards of Sarsa and Q-learning, and explain your result.(20%)



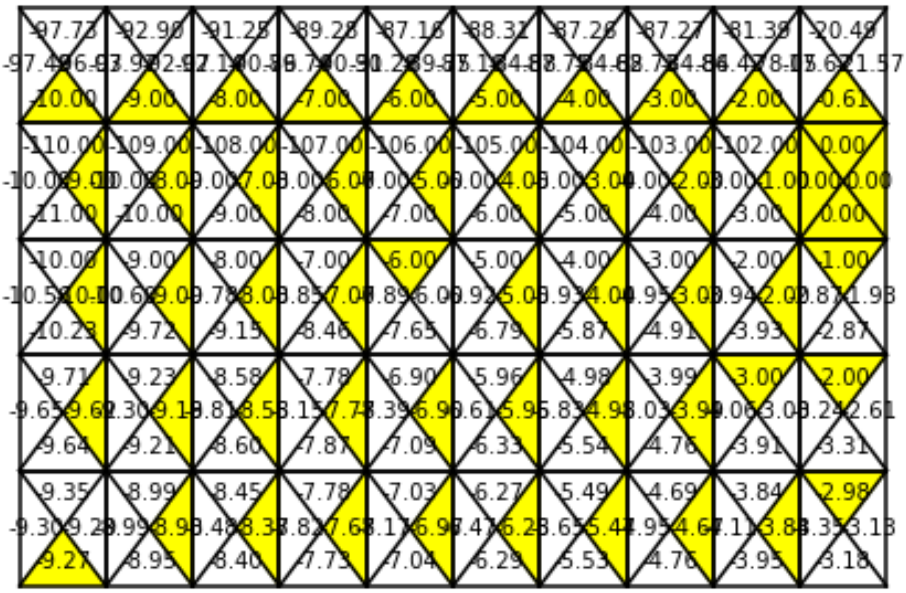




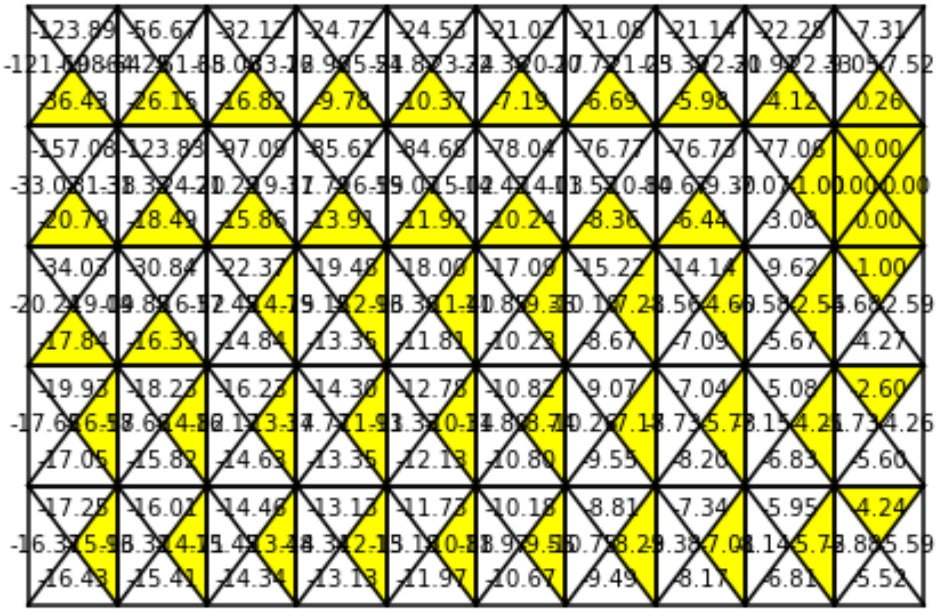
Sarsa 比 Q-learning 有比較好的線上效能，主要是因為 Sarsa 將動作選擇造成的影響考慮進 Q_table 的更新。

2. Plot the Q-values of Sarsa and Q-learning, and explain your result.(10%)

Q-learning Q_value

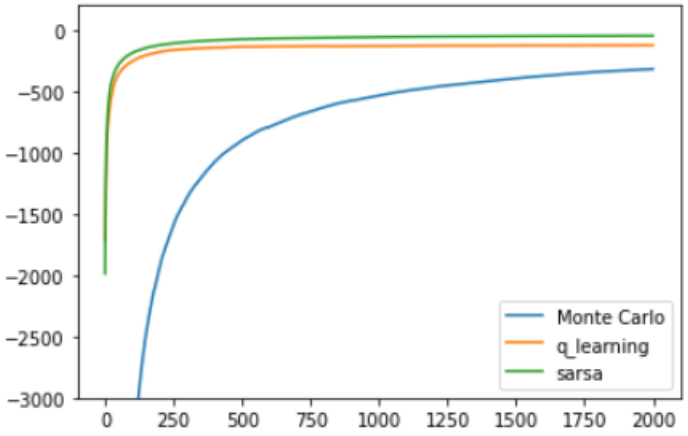


Sarsa Q_value



Q-learning 可以選擇最佳策略，直接從原點一直往右走到終點，因此代理人有時候會因為隨機動作選擇而走到沼澤裡。Sarsa 會得到次佳策略，選擇先往下繞道再往右後最後再往上到終點，避免因為隨機動作選擇進入沼澤。若 epsilon 值越大，Sarsa 選擇路徑會更保守。

3.Complete Monte Carlo, and compare average rewards.(10%)



Monte Carlo 演算法是離線學習，而它的代理人在執行任務過程中不做任何更新和學習，造成比較慢的學習速度相較 Q-learning 和 Sarsa，且需要大量記憶體儲存軌跡資料，而他如果任務需要較長時間執行的話，很難讓代理人轉移到終點狀態，因此它的結果表現不佳。