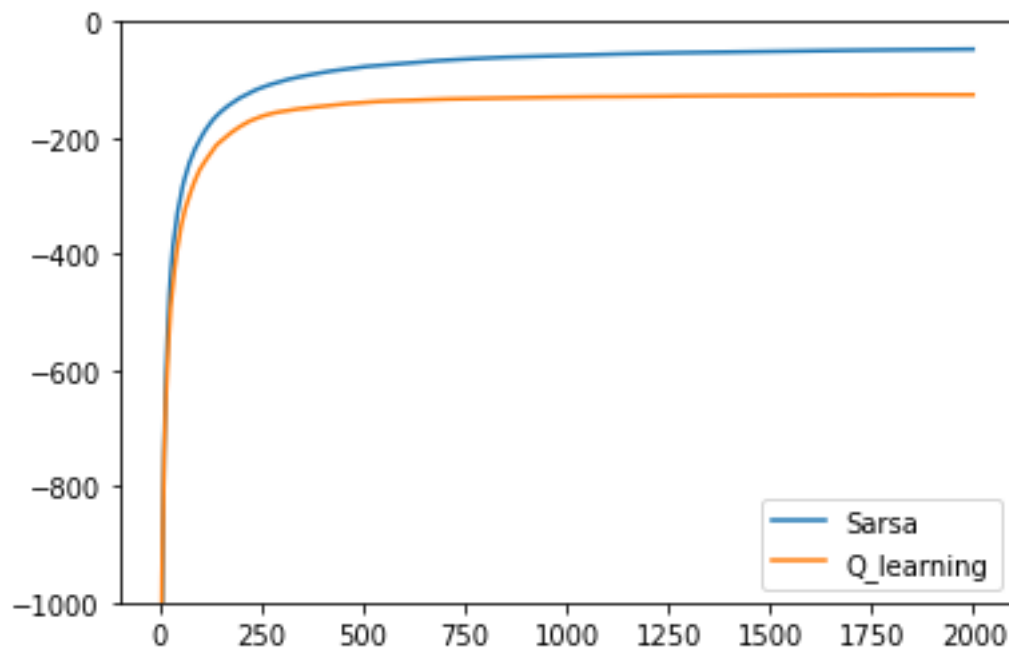


## Experiments and Analysis(40%)

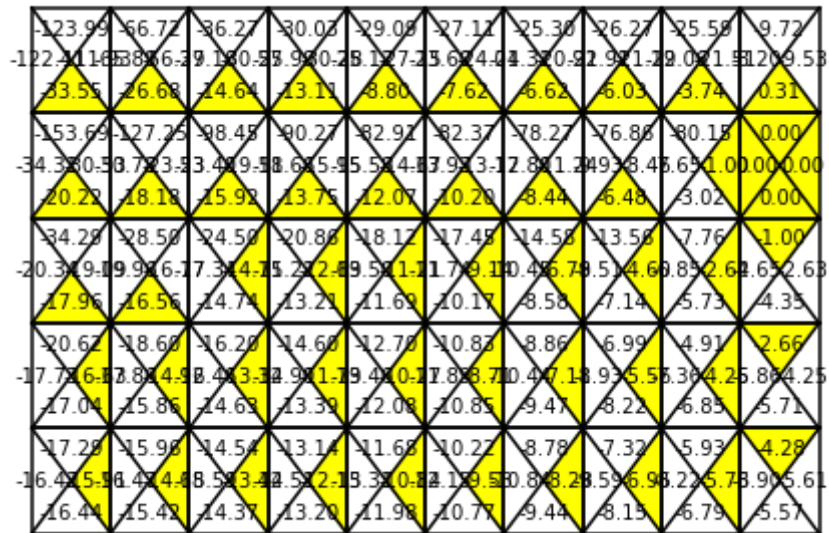
1. Plot the average rewards of Sarsa and Q-learning, and explain your result.(20%)



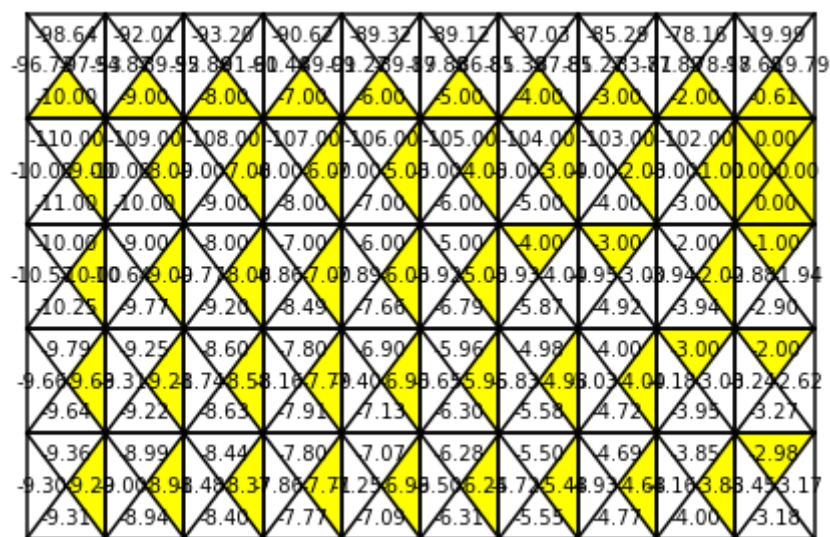
Sarsa 是 on-policy，會走最遠離 Swamp 的路徑，所以進入 Swamp 的機會較小，average rewards 就較大；反之，Q-learning 是 off-policy，會走最短路徑，所以進入 Swamp 的機會較大，average rewards 就較小。

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

Sarsa :

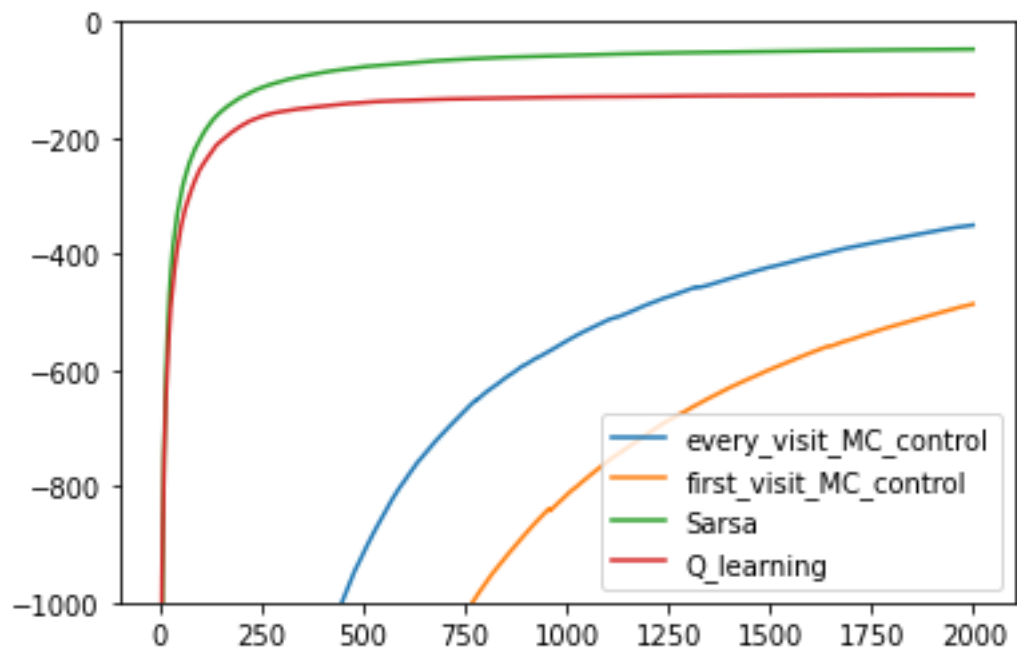


Q-learning :



Sarsa 是 on-policy，會走最遠離 Swamp 的路徑；反之，Q-learning 是 off-policy，會走最短路徑。

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



Sarsa 和 Q-learning 是每個 step 都更新，而蒙地卡羅是跑完整個 trajectory 才更新，所以 Sarsa 和 Q-learning 會學習得比蒙地卡羅快。