HW9 108061217 鍾永桓

result

從圖中的結果來看，actor_critic 比 sarsa 更快達到收斂，並且更為穩定，actor_critic 能收斂如此快的原因應該是因為它的報酬減去了基線並且它不用完整報酬的使用，所以先對於其他策略梯度更新更快收斂，而它跟 sarsa 的差別在於 actor_critic 是直接學習策略，而 sarsa 是學習價值函數，所以如果是龐大的動作空間，actor_critic 將會處理的比 sarsa 更好。