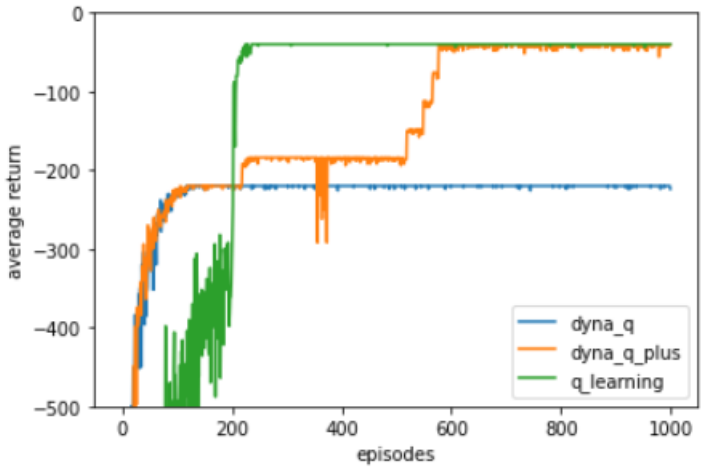
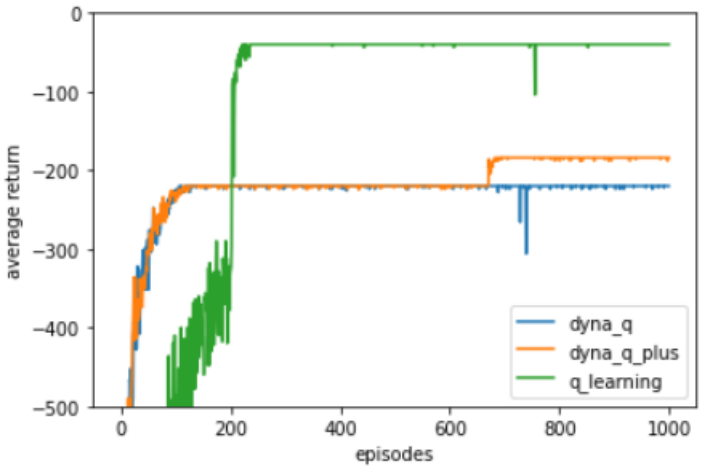
# Experiment 1 (20%)

In dyna-Q+ algorithm line20, τ need to restart in every episode or restart in every simulation? please do the experiment and explain your answer.

**根據實驗結果圖表顯示，若在每次 episode 就重置 τ，則 dyna-Q+ 的 returns 曲線在 400~600 左右 episodes 時不會像每次 simulation 就重置 τ 升起，這是因為一次 episode 走過的軌跡數不夠多，演算法難以估測有哪些狀態動作配對是明顯許久未拜訪過的，因此那些真正許久未拜訪的狀態無法有效地透過 kappa 來提升 reward，造成 agent 在學習時會沒有發現環境已發生改變而改善動作選擇策略**



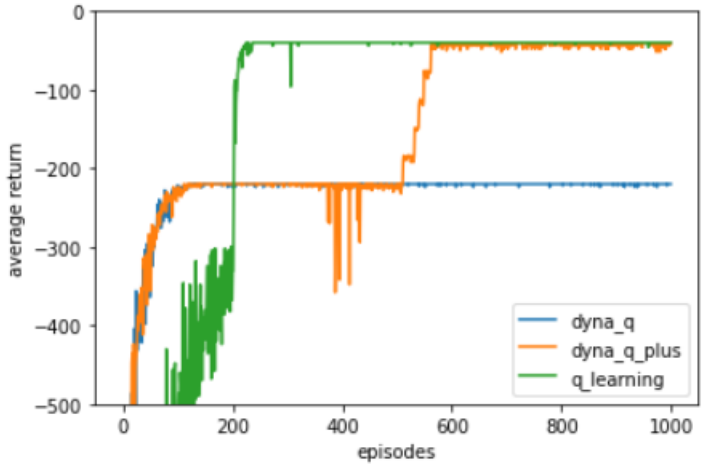

◆Restart τ in every simulation




◆Restart τ in every episode

# Result(60%)

```
dyna_q
dyna_q_plus
q_learning
```




# Question 1(20%)

Why Q-learning can react instantly when environment change?

因為 Q-learning 的特點是會大膽地隨機拜訪狀態，可以接受失敗的機會，所以可以找到最佳策略，在探索和開發上比 dynaQ 來的好，因此在像這樣環境會改變中，會立即地做出反應。

而 dynaQ 的目的是為了加速 Q-learning 的學習速度，因此引入了模型學習的機制，儲存已拜訪過的狀態動作配對於模型中，會倚賴模型中的資料學習，因此使用 dynaQ 會讓代理人只注意模型中儲存的狀態動作，而沒有察覺到牆壁開了一個洞後多出來的新路徑，而導致沒有比 Q-learning 更快速的反應。