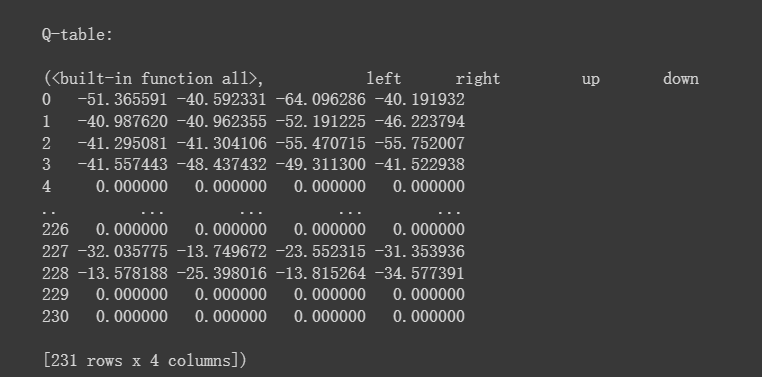
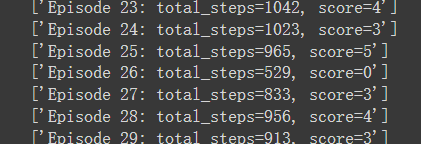
**HW3\_109403021**

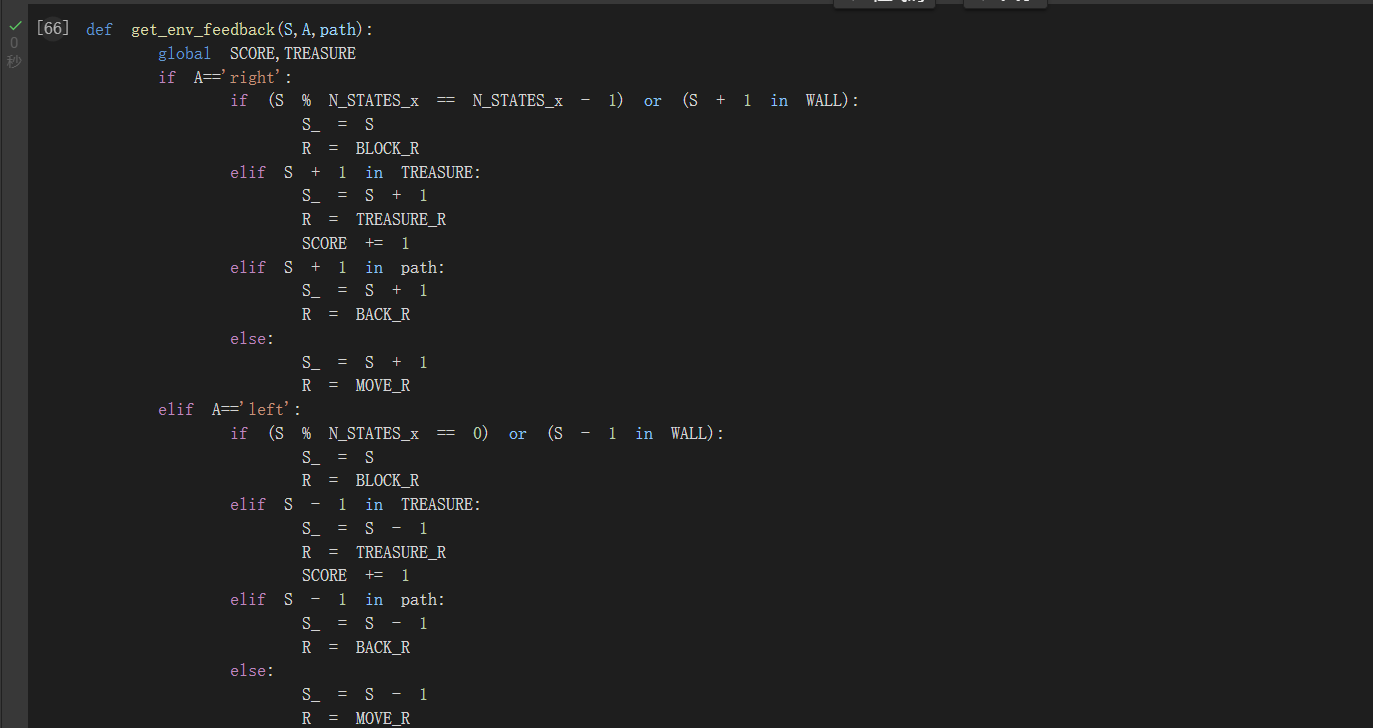
**Colab連結:** [**連結**](https://colab.research.google.com/drive/1TLP5yoEfaMQyk3-yYLPH1IWnw_Kpl920?usp=sharing)

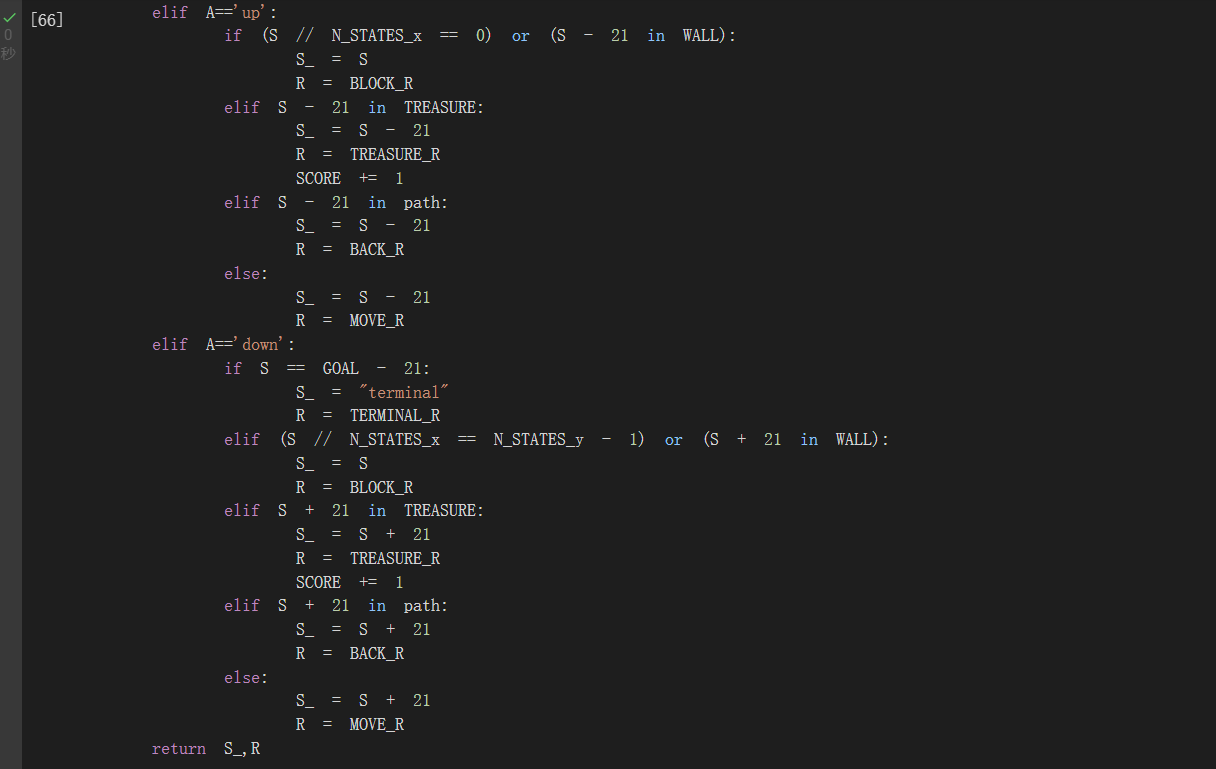
1. **訓練完的 Q-table**
2. **步數最少且寶藏數最多的截圖**

**Episode25: total\_steps=965, socre=5**

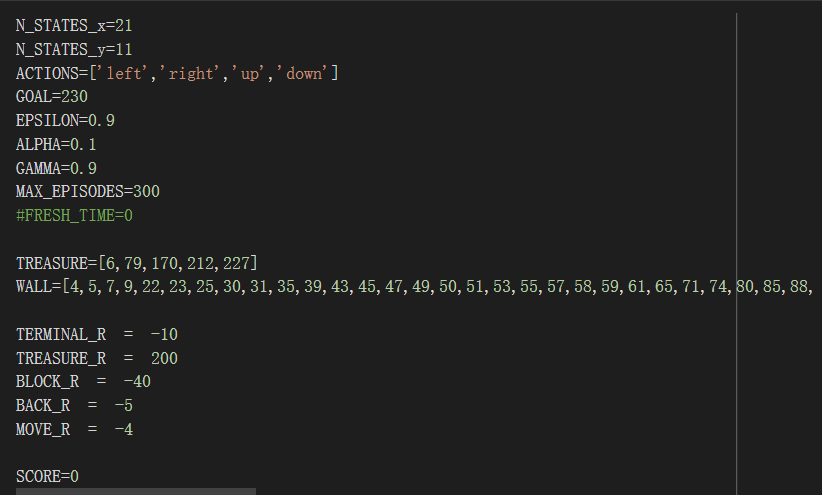


1. **Reward和參數設定**





依序為右、左、上、下的action內都給予了不同情況的下一步與reward指示，各別都依序為遇到邊界或牆壁就留在原地並reward為BLOCK\_R、遇到寶藏就移動到寶藏並reward為TREASURE\_R、遇到走過的路或新的空個就移動過去並分別給予reward為BACK\_R與MOVE\_R。而下的action中還給予了遇到終點並移動到終點給予reward為TERMINAL\_R的情況



參數設定上greedy rate、learing rate和discount rate我都有嘗試過使用其他數值，但效果沒有比叫好甚至常常更差，所以最後還是用回原本的。

Reward的參數設定上終點的設為負值可以阻止寶藏都沒拿就過早到達終點，寶藏重要性很高所以reward設比較大，撞牆reward自然也要設為負值。

然後正常移動MOVE\_R我本來試過0或1等等，但最後發現設為負的可以減少total\_steps而設為負數。而BACK\_R的部分我發現倍率和MOVE\_R相差越大容易過早收斂到不撿寶藏直接去終點，所以我的MOVE\_R和BACK\_R從[-1:-2]、[-2:-3]、[-3:-4]試到最後設為-5與-4。

1. **心得**

這次的作業我覺得是三次裡最有趣的，感覺很像在玩遊戲，調參數上感覺

有點像在幫遊戲角色配點以達到最佳，相對之前建模型更有感覺。

Reward 的if,else判斷那邊就差不多那樣不太花時間，主要還是花了很多時間調參數。我一開始還忘了在更新Q table那邊把每輪episode的score歸零，害我看到score一直是5步數還很少還白開心。後來發現不合理改掉之後才發現原來是一下子就收斂到只去終點score為0步數才少的情況，才發現不能像我本來想的終點最重要所以reward設最高、寶藏低一點這樣，甚至要把終點的reward設為負數避免沒撿到寶藏就過去，畢竟寶藏太重要了。

我也試過觀察地圖進去寶藏的路，讓上下左右一般移動獲得的reward不一樣促進拿寶藏，但寶藏有5個而且路很多所以不太可行的樣子。

我甚至還想過把貪婪度設很低讓它都隨機行動拚拚運氣，後來發現根本不實際，還是乖乖做好Q table更新才是對的。

雖然我最後做的蠻差的，結果還是會讓它收斂到步數少但score=0的情況，感覺還靠了一點運氣在前面的episode找到一個還可以的，但是過程確實蠻有趣，比起之前作業更會讓我想有空再試試怎麼讓結果更好甚至激發我更多研究機器學習的興趣，。