1. (10%) Policy Gradient 方法
   1. 請閱讀及跑過範例程式，並試著改進 reward 計算的方式。
   2. 請說明你如何改進 reward 的算法，而不同的算法又如何影響訓練結果？
2. (15%) 試著修改與比較至少三項超參數（神經網路大小、一個 batch 中的回合數等），並說明你觀察到什麼。
3. (15%) 請同學們從 Q Learning、Actor-Critic、PPO、 DDPG、TD3 等眾多 RL 方法中擇一實作，並說明你的實作細節。
4. (10%) 請具體比較（數據、作圖等）你實作的方法與 Policy Gradient 方法有何差異，並說明其各自的優缺點為何。