Data Science Hw4

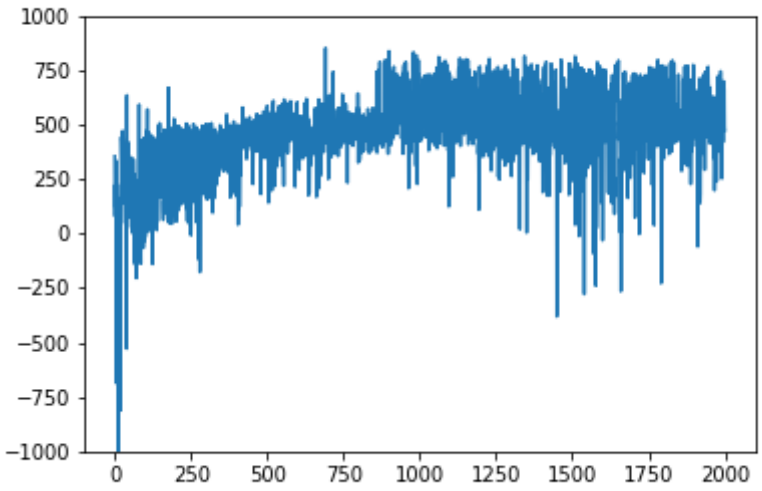
Algo

- 使用DQN algo
- ReplayBuffe紀錄agent得到的reward, state
- 使用兩個NN, policy與target, policy會一直更新,而tatget每玩20次才更新一次
- loss function使用mean square
- NN 架構為1 hidden layer, unit數目為50, activate function使用relu

reward shaping

1. 構想
    - 因為lander起始位置為(0, y'), 所以只要偏離x=0就給出懲罰
      reward = reward - abs(state[0])
    - 觀察到如果什麼都不調整,則lander只會一直待在y=y'之處, 所以給出y座標越高會有懲罰
      reward = reward - abs(state[1])
    - 因為旋轉會造成lander的不平穩, 所以只有旋轉就給出懲罰
      reward = reward - abs(state[4])
    - 要鼓勵lander往下降落,與懲罰往上升,所以給出
      reward = reward - state[3] *10
    - 與此同時y軸的速度會與y軸的座標相互制衡,達到越接近y=0時,lander降落的速度會趨於平緩

2. 實驗結果



3. gif