

HW5

學號:107062631

姓名:方晟軒

1. TODO

演算法實作 TODO 的部分基本上只要把

$\text{reward} += \text{abs}(\text{next_state}[0] + 0.6) + \text{next_state}[1] * 8$ 這行調整參數即可完成，不過我有試著去做 reward shaping 的部分，下面這張圖是根據高度所調整的 reward，如果前一次的 state 高度小於現在 state 的高度，則要增加他的 reward，反之則減少，讓他有目的的去學習

```
#reward += abs(next_state[0]+0.8) + next_state[1]*10
if state[0] <= next_state[0]:
    reward += abs(next_state[0]+0.8) + next_state[1]*10
    #hint: abs(next_state[0]+c) + next_state[1]*k
    #reward += abs(next_state[0]+0.6) + next_state[1]*8
else:
    reward -= 0.01
```

而這個是根據速度來決定，速度越大表示能到達的高度越高，用這個來決定第二種 reward shaping 的方法

```
#reward += abs(next_state[0]+0.8) + next_state[1]*10
if state[1] < next_state[1]:
    reward += abs(next_state[0]+0.8) + next_state[1]*10
    #hint: abs(next_state[0]+c) + next_state[1]*k
    #reward += abs(next_state[0]+0.6) + next_state[1]*8
else:
    reward -= 0.01
```