

EE473 Deep Reinforcement Learning Homework 1

Author: Allen Liu

Exercise 3.1

- The deep reinforcement learning can be applied to how to drive along a road.
 - *State*: The velocity, position, steering of the vehicle.
 - *Action*: The power applied to the engine, the force applied to the steering.
 - *Reward*: The reward will +1 when it passes a checkpoint.
- The deep reinforcement learning can be applied to how to invest money
 - *State*: State is how much money have in the account.
 - *Action*: How much money to invest to a company.
 - *Reward*: Number of profits gained from the investment.
- The deep reinforcement learning can be applied to how to maintain proper room temperature.
 - *State*: The room temperature
 - *Action*: Power applied to the state and the heater and AC.
 - *Reward*: The reward will +1 when it reaches the desired one.

Exercise 3.3

The place to draw the line between the agent and the environment is by choosing the action \mathcal{A}_1 to be acceleration, steering wheel and the break. In this case, the agent is the human body, where the environment is the vehicle and its surroundings.

The basis to choose the line should based on what state parameters are easily to be measurable. For example, for the action \mathcal{A}_1 , the state \mathcal{S}_1 will be vehicle's speed and heading. In another hand, if the action \mathcal{A}_2 were chosen to be the tire torque, the environment will be the ground and the surrounding vehicles, and the agent will be the vehicle and the human body. Thus state \mathcal{S}_2 will be force distribution over the ground. This is difficult to be measured.

The reason to choose one over another is to ensure that states S and actions A always has a smallest set of parameters and are easily measurable.

Exercise 4.4

s	a	s'	$p(s', r s, a)$	$r(s, a, s')$
high	search	high	α	r_{search}
high	search	low	$1 - \alpha$	r_{search}
low	search	high	β	-3
low	search	low	$1 - \beta$	r_{search}
high	wait	high	1	r_{wait}
low	wait	low	1	r_{wait}
low	recharge	high	1	0

Investing in Stocks

Part 1

The total return value is 395.42 dollars.

```
import random
import matplotlib.pyplot as plt

if __name__ == "__main__":
    stock_val = 100.0

    values = []
```

```

for i in range(100):
    val_s1 = stock_val / 2
    val_s2 = stock_val / 2

    ret_s1 = random.uniform(-1.0, 1.0)
    ret_s2 = random.uniform(-2.0, 8.0)

    new_s1 = val_s1 * (1.0 + ret_s1 / 100.0)
    new_s2 = val_s2 * (1.0 + ret_s2 / 100.0)

    stock_val = new_s1 + new_s2

    print(f"Total value at day {i + 1} is {stock_val}")

    values.append(stock_val)

plt.plot(values)
plt.xlabel("Days")
plt.ylabel("Total Return [$]")
plt.show()

```

Code output:

```

jliu@nu-msr-jliu:~/Course_Materials/ELEC_ENG_473/Homeworks/Homework1 (main)$ python stock.py
Total value at day 1 is 100.04639547788108
Total value at day 2 is 102.93861386699601
Total value at day 3 is 104.64591052165072
Total value at day 4 is 105.44776989256496
Total value at day 5 is 106.99662024549254
Total value at day 6 is 110.22649237042509
Total value at day 7 is 110.01431273678641
Total value at day 8 is 113.46070340816183
Total value at day 9 is 116.85674522138338
Total value at day 10 is 120.75784502334403
Total value at day 11 is 120.83497648951985
Total value at day 12 is 124.06695207891931
Total value at day 13 is 123.29276177364966
Total value at day 14 is 126.4062532253925
Total value at day 15 is 128.0303285925138
Total value at day 16 is 130.84692293649044
Total value at day 17 is 131.1332955061593
Total value at day 18 is 132.57889326752098
Total value at day 19 is 137.52147503060914
Total value at day 20 is 140.56213837371425
Total value at day 21 is 141.0935325207784
Total value at day 22 is 142.987211946043
Total value at day 23 is 144.214142784861
Total value at day 24 is 145.09635436117773
Total value at day 25 is 145.13172171491425
Total value at day 26 is 144.8683655118191
Total value at day 27 is 149.57886041068065
Total value at day 28 is 148.45937953778662
Total value at day 29 is 148.25216748965119
Total value at day 30 is 152.273740422845
Total value at day 31 is 155.57224192745426
Total value at day 32 is 157.38241138486399
Total value at day 33 is 162.64206757479135

```

Total value at day 34 is 161.92854967730216
Total value at day 35 is 166.12520901034227
Total value at day 36 is 167.3578805025246
Total value at day 37 is 171.82343441765357
Total value at day 38 is 177.55726883321262
Total value at day 39 is 182.77241047250612
Total value at day 40 is 184.93178796948584
Total value at day 41 is 183.8035853863604
Total value at day 42 is 186.20020488287875
Total value at day 43 is 185.05752289087735
Total value at day 44 is 188.66807416968186
Total value at day 45 is 193.96460988877223
Total value at day 46 is 199.25106693985126
Total value at day 47 is 206.2819085972173
Total value at day 48 is 210.77854707541854
Total value at day 49 is 212.03177564185324
Total value at day 50 is 209.09093229464736
Total value at day 51 is 211.78430028826554
Total value at day 52 is 213.53909981586608
Total value at day 53 is 219.52868432118822
Total value at day 54 is 223.73097515557922
Total value at day 55 is 226.8376268556359
Total value at day 56 is 232.79242920502398
Total value at day 57 is 235.7510513563678
Total value at day 58 is 239.4784985544089
Total value at day 59 is 247.19012008155363
Total value at day 60 is 250.9893408973175
Total value at day 61 is 249.24680292541007
Total value at day 62 is 247.69020918278582
Total value at day 63 is 248.55062959410537
Total value at day 64 is 255.60805467796473
Total value at day 65 is 257.337476803893
Total value at day 66 is 258.37037509367144
Total value at day 67 is 257.6387524909891
Total value at day 68 is 257.84463108721
Total value at day 69 is 255.77848183960276
Total value at day 70 is 256.86795357890395
Total value at day 71 is 263.9588301761596
Total value at day 72 is 270.0835811105806
Total value at day 73 is 270.202292372188
Total value at day 74 is 272.121814899702
Total value at day 75 is 272.3285086836138
Total value at day 76 is 278.3447818963665
Total value at day 77 is 284.7763673286778
Total value at day 78 is 287.5974442927135
Total value at day 79 is 285.2431099041612
Total value at day 80 is 296.3724623208166
Total value at day 81 is 308.05381427247835
Total value at day 82 is 313.36282568262203
Total value at day 83 is 318.6660230666808
Total value at day 84 is 326.93416689826415
Total value at day 85 is 341.0122672990381
Total value at day 86 is 342.9484682688682
Total value at day 87 is 342.2726627317122
Total value at day 88 is 345.2913061463556
Total value at day 89 is 346.20532600194565

Total value at day 90 is 356.1697469174444
Total value at day 91 is 360.6542542246973
Total value at day 92 is 372.8225964324525
Total value at day 93 is 370.81839853538963
Total value at day 94 is 373.4696595660231
Total value at day 95 is 385.119453500404
Total value at day 96 is 387.26346884949265
Total value at day 97 is 384.61627901505176
Total value at day 98 is 395.29821750634267
Total value at day 99 is 397.77212414944665
Total value at day 100 is 395.4154338142787

Part 2

- *State \mathcal{S}* : Total net return value
- *Action \mathcal{A}* : The percentage to invest on each stock.
- States \mathcal{S} can be any value greater than 0, actions \mathcal{A} can range from 0 to 1 for both stocks, with total not exceeding 1.
- Rewards is the number of money gained each day, can be any value.

Part 3

Since stock 1 have a expectation of 0% while stock 2 have a expectation of 3%. So it is better to invest on stock 2.