Report

Name: Chen Xingyi UID: 3036198102

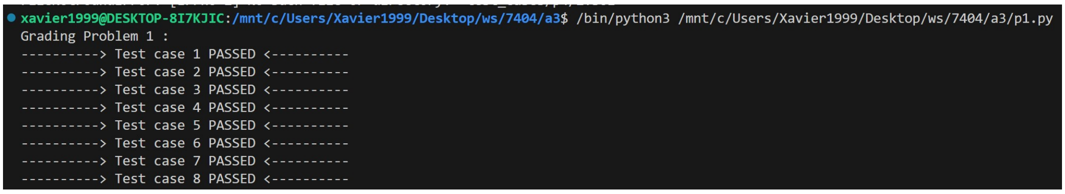
Parser:

Completed (Runtime < 1s).

P1:

Completed. (Runtime < 1s).

Challenge: Hardly any challenge.

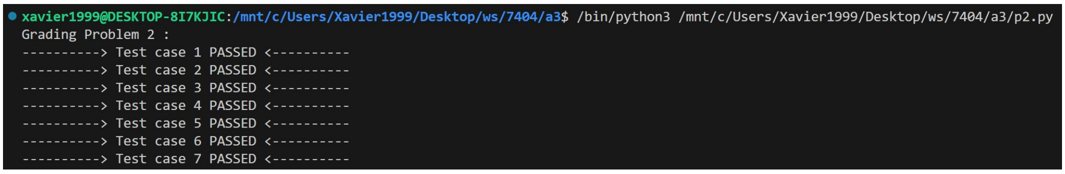


```
● xavier1999@DESKTOP-8I7KJIC:/mnt/c/Users/Xavier1999/Desktop/ws/7404/a3$ /bin/python3 /mnt/c/Users/Xavier1999/Desktop/ws/7404/a3/p1.py
  Grading Problem 1 :
  ----------> Test case 1 PASSED <----------
  ----------> Test case 2 PASSED <----------
  ----------> Test case 3 PASSED <----------
  ----------> Test case 4 PASSED <----------
  ----------> Test case 5 PASSED <----------
  ----------> Test case 6 PASSED <----------
  ----------> Test case 7 PASSED <----------
  ----------> Test case 8 PASSED <----------
```

P2:

Completed. (Runtime < 1s).

Challenge: Hardly any challenge.

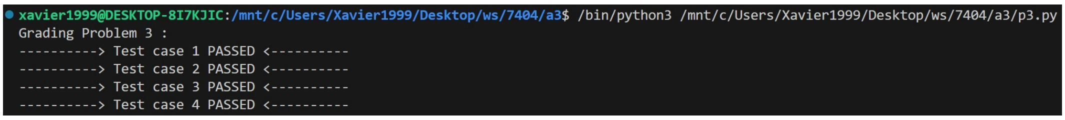


```
● xavier1999@DESKTOP-8I7KJIC:/mnt/c/Users/Xavier1999/Desktop/ws/7404/a3$ /bin/python3 /mnt/c/Users/Xavier1999/Desktop/ws/7404/a3/p2.py
  Grading Problem 2 :
  ----------> Test case 1 PASSED <----------
  ----------> Test case 2 PASSED <----------
  ----------> Test case 3 PASSED <----------
  ----------> Test case 4 PASSED <----------
  ----------> Test case 5 PASSED <----------
  ----------> Test case 6 PASSED <----------
  ----------> Test case 7 PASSED <----------
```

P3:

Completed. (Runtime < 1s).

Challenge: Hardly any challenge.



```
● xavier1999@DESKTOP-8I7KJIC:/mnt/c/Users/Xavier1999/Desktop/ws/7404/a3$ /bin/python3 /mnt/c/Users/Xavier1999/Desktop/ws/7404/a3/p3.py
  Grading Problem 3 :
  ----------> Test case 1 PASSED <----------
  ----------> Test case 2 PASSED <----------
  ----------> Test case 3 PASSED <----------
  ----------> Test case 4 PASSED <----------
```

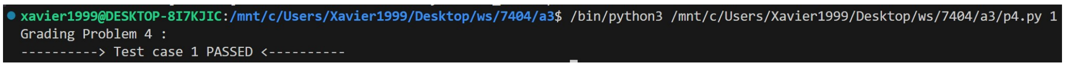
P4:

Completed. (Runtime < 1s).

Challenge:

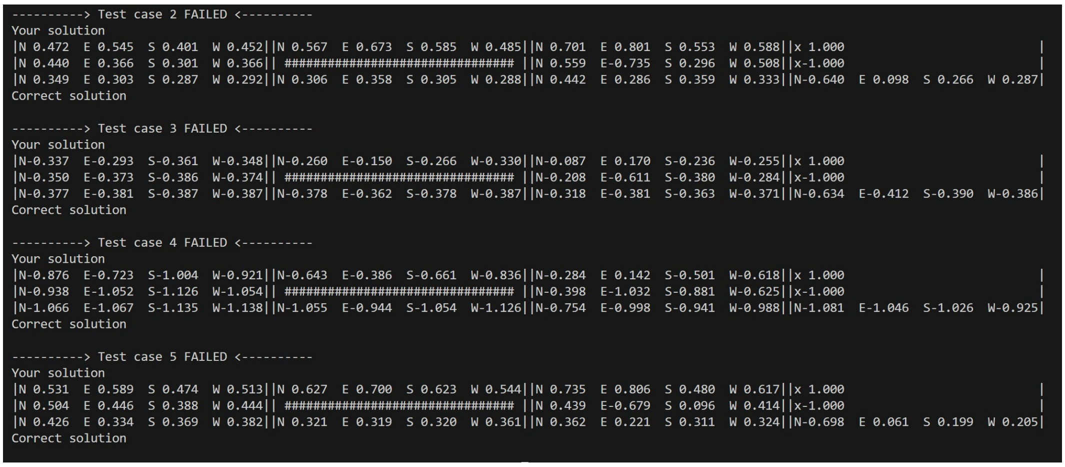
My algorithm successfully identifies the optimal policy approximately 65% of the time. However, upon closer examination, I noticed that deviations from the optimal policy tend to occur when the 3rd column of the 3rd row contains the action N instead of W.

Analysis: The random directions selected during the learning process can significantly influence the resulting optimal policy. Considering that each action incurs a negative living reward, if a particular direction is chosen more frequently, its associated Q value tends to decrease. This effect is particularly pronounced at the outset when the learning rate is high. The cumulative impact of living rewards may lead to a substantial reduction in the Q value for that specific direction. Consequently, the algorithm might favor an alternative direction as the optimal policy for a given state if its Q value surpasses that of the originally preferred direction. Q-Value Impact: The living reward’ s influence can cause a substantial reduction in the Q-value for a specific direction. Consequently, if the Q-value for the optimal policy’ s direction is smaller than that for other actions, the algorithm may choose an alternative direction as the best policy for a given state.



```
● xavier1999@DESKTOP-8I7KJIC:/mnt/c/Users/Xavier1999/Desktop/ws/7404/a3$ /bin/python3 /mnt/c/Users/Xavier1999/Desktop/ws/7404/a3/p4.py 1
  Grading Problem 4 :
  ----------> Test case 1 PASSED <----------
```

The converged policy and the converged Q values for the left test cases in p4:



```
----------> Test case 2 FAILED <----------
Your solution
|N 0.472  E 0.545  S 0.401  W 0.452||N 0.567  E 0.673  S 0.585  W 0.485||N 0.701  E 0.801  S 0.553  W 0.588||x 1.000                          |
|N 0.440  E 0.366  S 0.301  W 0.366|| ############################### ||N 0.559  E-0.735  S 0.296  W 0.508||x-1.000                          |
|N 0.349  E 0.303  S 0.287  W 0.292||N 0.306  E 0.358  S 0.305  W 0.288||N 0.442  E 0.286  S 0.359  W 0.333||N-0.640  E 0.098  S 0.266  W 0.287|
Correct solution

----------> Test case 3 FAILED <----------
Your solution
|N-0.337  E-0.293  S-0.361  W-0.348||N-0.260  E-0.150  S-0.266  W-0.330||N-0.087  E 0.170  S-0.236  W-0.255||x 1.000                          |
|N-0.350  E-0.373  S-0.386  W-0.374|| ############################### ||N-0.208  E-0.611  S-0.380  W-0.284||x-1.000                          |
|N-0.377  E-0.381  S-0.387  W-0.387||N-0.378  E-0.362  S-0.378  W-0.387||N-0.318  E-0.381  S-0.363  W-0.371||N-0.634  E-0.412  S-0.390  W-0.386|
Correct solution

----------> Test case 4 FAILED <----------
Your solution
|N-0.876  E-0.723  S-1.004  W-0.921||N-0.643  E-0.386  S-0.661  W-0.836||N-0.284  E 0.142  S-0.501  W-0.618||x 1.000                          |
|N-0.938  E-1.052  S-1.126  W-1.054|| ############################### ||N-0.398  E-1.032  S-0.881  W-0.625||x-1.000                          |
|N-1.066  E-1.067  S-1.135  W-1.138||N-1.055  E-0.944  S-1.054  W-1.126||N-0.754  E-0.998  S-0.941  W-0.988||N-1.081  E-1.046  S-1.026  W-0.925|
Correct solution

----------> Test case 5 FAILED <----------
Your solution
|N 0.531  E 0.589  S 0.474  W 0.513||N 0.627  E 0.700  S 0.623  W 0.544||N 0.735  E 0.806  S 0.480  W 0.617||x 1.000                          |
|N 0.504  E 0.446  S 0.388  W 0.444|| ############################### ||N 0.439  E-0.679  S 0.096  W 0.414||x-1.000                          |
|N 0.426  E 0.334  S 0.369  W 0.382||N 0.321  E 0.319  S 0.320  W 0.361||N 0.362  E 0.221  S 0.311  W 0.324||N-0.698  E 0.061  S 0.199  W 0.205|
Correct solution
```

approximate number of hours: 9 hours