* Exercise 0 *

Q.3) How does manual policy compares to random policy?
What are some reasons for difference in performance?

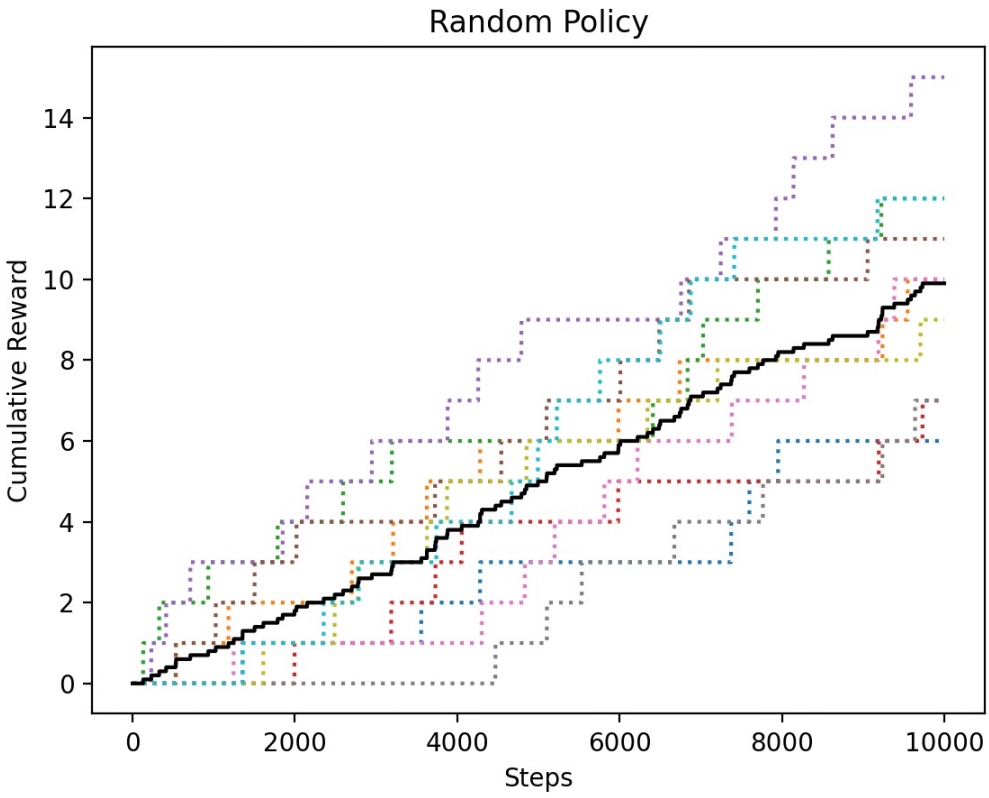
→
- Generally,
  Manual Policy performs Better than random policy & agent will reach to goal state faster than random policy.
- Some reasons are as follows:-
  i) When we manually instruct our agent to reach goal state, we can make sure that agent follows a close to optimal path, as we can observe entire environment. In doing so, we eliminate trial & error.

ii) Whereas, in random policy agent will explore each step, with equal probability. For large number of iterations, only 1 in 4 steps taken would lead to goal state.

iii) Keeping in mind that we added some noise (0.1 probability of going to each of perpendicular directions)

But even if we consider this noise, agent will go in the right direction (provided by human) with 0.8 probability. which is far better than 0.25 probability of random policy.

iv) Above comparison assumes that human guided manual policy will try to reach goal state in minimum steps possible.


Random Policy
Cumulative Reward
Steps
0
2
4
6
8
10
12
14
0
2000
4000
6000
8000
10000

Q.4) Describe strategy used by better & worse policies & why it leads to generally worse/better performance?

- Better policy :-

Strategy :- Assign higher probabilities to UP & RIGHT directions.

Why it works? :-

- As we know, goal state lies in (10x10) location of matrix & agent always starts from (0x0) location.

- Thus intuitively, agent travels from Bottom left → Top Right

- we can use this prior knowledge & feed it to actions taken by agent.
  So, we assigned
  35% chance of going UP
  35% chance of going RIGHT
  15% chance of going DOWN
  15% chance of going LEFT

- This ultimately results in higher AVG cumulative rewards than Random policy.

- Additionally we can add some in between rewards to guide agent to the goal state & use greedy policy on rewards instead of actions.

- Worse policy :-

Strategy :- Perform opposite operation than better policy.

Assign slightly higher probabilities for going DOWN & LEFT.

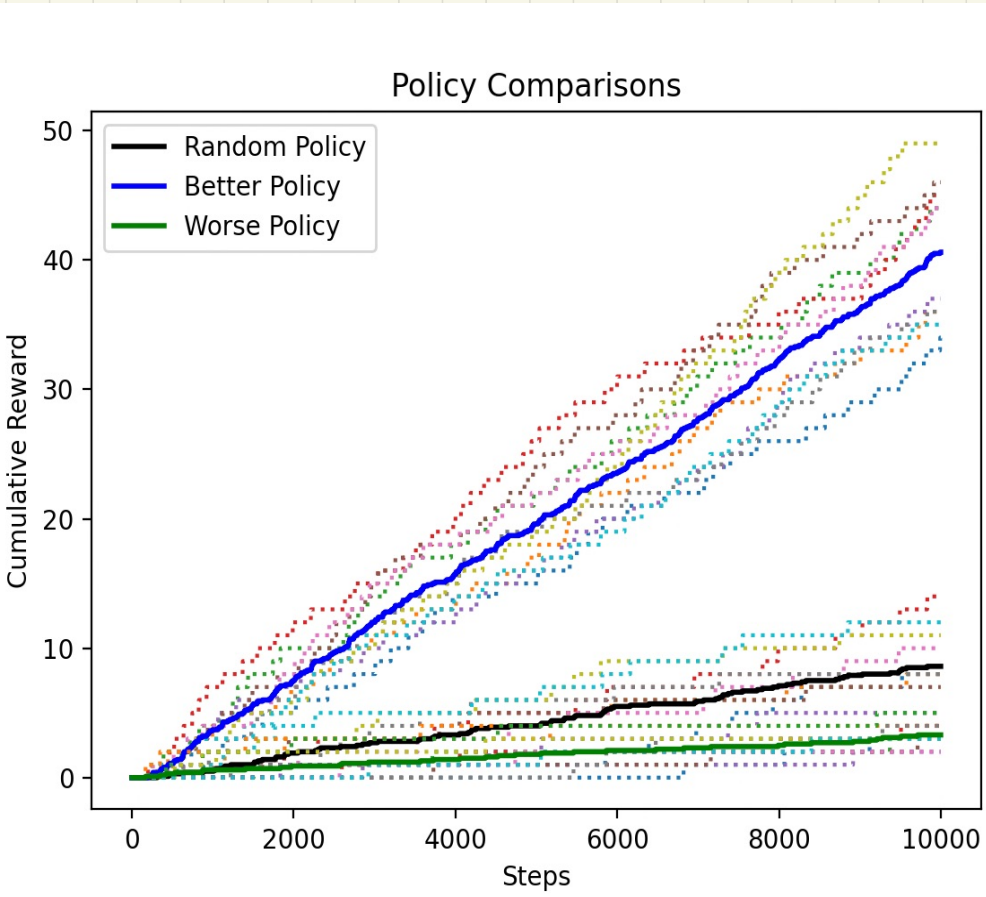
which will keep agent from reaching goal state easily.

Why it works? :-

- we use the same context for environment as we used in better policy & make sure agent hardly takes steps toward goal.
- Here we assigned slightly higher probabilities to go DOWN & LEFT.

DOWN	27.5%
LEFT	27.5%
UP	22.5%
RIGHT	22.5%

- we only changed probabilities slightly because otherwise agent would only get stuck at 0 reward for a long time
- Instead of this, we can also have agent move to only single or two directions which will stop it from reaching goal state.


Policy Comparisons
Random Policy
Better Policy
Worse Policy
Cumulative Reward
Steps
0
10
20
30
40
50
0
2000
4000
6000
8000
10000