CSE 4/546: Reinforcement Learning
Spring 2025

Instructor: Alina Vereshchaka

Assignment 1 - Defining & Solving RL Environments

 # References

- https://gymnasium.farama.org/environments/toy_text/
- https://gymnasium.farama.org/tutorials/gymnasium_basics/environment_creation/#sphx-glr-tutorials-gymnasium-basics-environment-creation-py
- https://cs.stanford.edu/people/karpathy/reinforcejs/

Part 1: Defining RL Environments [30 points]

Describe the deterministic and stochastic environments, including their sets of actions, states, rewards, main objectives, etc.

For this assignment, I have chosen an Autonomous Drone Delivery environment, where a drone must pick up two packages and deliver them to their respective destinations. There are two different scenarios in this environment:

Deterministic Environment

A deterministic environment means that the outcomes of actions are fixed and predictable. The same action in the same state always leads to the same result.

Challenges in the Deterministic Environment:

- The environment is fixed, meaning all elements (drone, packages, delivery locations, and obstacles) remain the same in every run.
- There is a tornado, which acts as a no-fly zone, meaning the drone cannot pass through it.
- The tornado is surrounded by wind that pushes the drone in a specific direction. If the drone enters a wind-affected cell, it has a higher probability of moving in the wind's direction.
- And two birds act as obstacles, each imposing a negative reward.
- Since the environment does not change dynamically, the drone can learn an optimal path and follow it every time without unexpected disturbances.

Stochastic Environment

A stochastic environment means that outcomes are random to some degree. The same action in the same state may lead to different results each time.

Challenges in the Stochastic Environment:

- Random Placement: Every time the environment is initialised, the drone, packages, delivery locations, birds, tornado, and wind zones appear in random positions on the grid.
- Dynamic Elements:
  - The tornado moves around the grid, changing its location at every time step.
  - The wind direction also shifts, making navigation harder.
  - There are birds, which might collide with the drone.
- Increased Difficulty: The drone must learn and adapt to different scenarios in each run instead of relying on a fixed strategy.
- This makes the task more complex, as the drone cannot follow a fixed path. Instead, it must learn to make decisions on the fly based on changing conditions.

Environment Setup

The environment is represented as a 6x6 grid that models a city layout where the drone performs deliveries.

Grid Properties:

- Size: 6x6 cells.
- Obstacles: No-fly zones (static obstacles) that the drone cannot cross.
- Goal: The drone must pick up both packages and deliver them to their respective destinations.
- Actions Available:

The drone can take the following actions:

- Up
- Down
- Left
- Right
- Pick up a package
- Drop off a package

Rewards System:
The drone earns or loses points based on its actions and performance.

Event	Reward/Penalty
Successfully delivering a package	100
Picking up a package	25
Entering a no-fly zone (tornado)	-100
Getting hit by a bird	-50
Being pushed by the wind	-10
Taking a step (movement cost, **not yet implemented,** will implement in part 2)	-1

Note: Some rewards are missing in the current checkpoint and will be added in Part 2 to finalise the reward system for optimisation. This includes penalties for failed drop-offs and failed pick-ups.

Terminal State
The drone successfully delivers both packages and earns maximum rewards.

Provide visualisations of your environments.

How did you define the stochastic environment?

In my stochastic environment, all objects are placed randomly at the start, and some elements continue to change as the drone moves. This creates an unpredictable environment, making it more challenging for the drone to complete its task.

Key Features of the Stochastic Environment:

Random Initialization:

At the beginning of each simulation, the positions of the drone, two packages, two delivery locations, tornado, wind zones, and birds are all placed randomly on the grid.
This ensures that the drone does not start in the same location every time and must adapt to different scenarios.

Dynamic Tornado and Wind Movement:

The tornado, which acts as a no-fly zone, moves randomly during the simulation.
The wind zones surrounding the tornado also change their position, increasing the difficulty of navigation.
If the drone enters a wind zone, it has a higher probability of being pushed in the wind's direction.

Real-Time Decision Making:

Because the tornado and wind move unpredictably, the drone cannot rely on a fixed path like in a deterministic environment.
Instead, it must constantly analyse the environment and make real-time decisions to avoid obstacles while still reaching the delivery destinations efficiently.

What is the difference between deterministic and stochastic environments?

Deterministic Environment:

- Everything in the environment is fixed and predictable.
- The drone, packages, delivery locations, tornado, and wind zones stay in the same positions every time the simulation starts.
- The tornado is a no-fly zone, and the wind always pushes the drone in the same direction.
- The drone can learn an optimal path and follow it without surprises.

Stochastic Environment:

- Everything in the environment is random and changes over time.
- The drone, packages, delivery locations, birds, tornado, and wind zones are placed randomly at the start.
- During the drone's movement, the tornado moves randomly with wind directions, making navigation harder.
- The drone must make real-time decisions instead of following a fixed path.

Main Difference:

In the deterministic environment, the outcome of every action is always the same, making it easier to plan a path.
In the stochastic environment, randomness makes each run different, forcing the drone to adapt to new challenges every time.

Safety in AI: Write a brief review (~5 sentences) explaining how you ensure the safety of your environments. E.g. how do you ensure that agent choose only actions that are allowed, that agent is navigating within defined state-space, etc.

To ensure the safety of my environment, I apply strict checks to prevent illegal actions and enforce correct behavior:

Boundary Checks: Before moving up, down, left, or right, the drone checks if the movement stays within the grid boundaries. If an action would move it outside, it is not allowed.

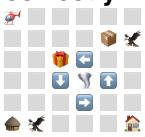
Valid Pick-Up Action: The drone can only pick up a package if it is in the same cell as the package. If it tries to pick up from an empty location, it receives a penalty.

Valid Drop-Off Action: The drone can only drop off a package if it is in the correct destination cell. If it drops a package at the wrong location, the package is moved to a new box, and the drone gets a negative reward.

No-Fly Zones & Obstacles: The drone is not allowed to enter tornado zones or other restricted areas. If it tries, it gets a large penalty.

By enforcing these rules, I ensure that the agent only takes legal actions, stays within the grid, and follows correct pick-up and drop-off procedures while learning to navigate efficiently. 🚁✅

Random Drone movement in deterministic env for 10+ steps to show that the environment logic is defined correctly.



```
Drone moved to (1, 0) Accumulated
Reward: 0
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
```

```
Attempted pickup failed
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Attempted pickup failed
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Drone moved to (2, 0) Accumulated
Reward: 0
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
```

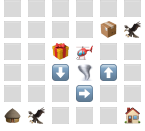
```
Drone moved to (3, 0) Accumulated
Reward: 0
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
```

```
Drone moved to (2, 0) Accumulated
Reward: 0
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
```

```
Attempted dropoff failed (no package
carried)
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Drone moved to (3, 0) Accumulated
Reward: 0
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
```

```
Drone moved to (3, 1) Accumulated
Reward: 0
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
```

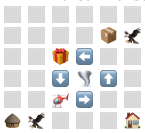
```
Drone moved to (3, 2) Accumulated
Reward: -10
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
```



```
Attempted pickup failed
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Drone moved to (3, 3) Accumulated
Reward: -110
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
```

```
Drone moved to (2, 3) Accumulated
Reward: -120
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
```

```
Attempted dropoff failed (no package
carried)
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Attempted pickup failed
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Drone moved to (2, 4) Accumulated
Reward: -120
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
```



```
Drone moved to (1, 4) Accumulated
Reward: -120
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
```

```
Attempted dropoff failed (no package
carried)
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Drone moved to (1, 5) Accumulated
Reward: -170
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
```

```
Drone moved to (1, 4) Accumulated
Reward: -170
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
```

```
Attempted dropoff failed (no package
carried)
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Attempted dropoff failed (no package
carried)
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Drone moved to (2, 4) Accumulated
Reward: -170
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
```

```
Drone moved to (3, 4) Accumulated
Reward: -180
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
```

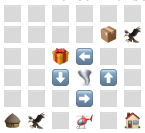
```
Drone moved to (3, 5) Accumulated
Reward: -180
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
```

```
Attempted dropoff failed (no package
carried)
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Attempted dropoff failed (no package
carried)
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Drone moved to (4, 5) Accumulated
Reward: -180
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
```

```
Drone moved to (3, 5) Accumulated
Reward: -180
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
```

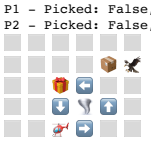
```
Drone moved to (3, 4) Accumulated
Reward: -190
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
```

```
Attempted pickup failed
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Drone moved to (3, 5) Accumulated
Reward: -190
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
```

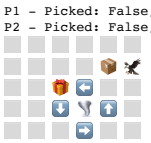


```
Attempted dropoff failed (no package
carried)
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Drone moved to (2, 5) Accumulated
Reward: -190
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
```

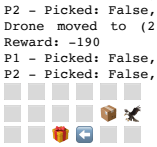
```
Attempted pickup failed
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Drone moved to (2, 4) Accumulated
Reward: -190
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
```



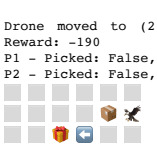
```
Attempted pickup failed
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Drone moved to (2, 5) Accumulated
Reward: -190
```



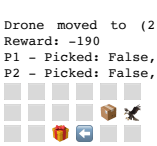
```
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
```



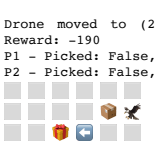
```
Attempted pickup failed
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Drone moved to (2, 4) Accumulated
Reward: -190
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
```



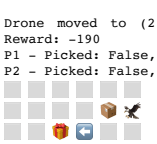
```
Drone moved to (2, 5) Accumulated
Reward: -190
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
```



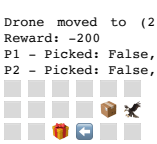
```
Drone moved to (2, 4) Accumulated
Reward: -190
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
```



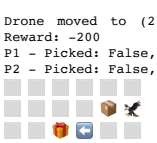
```
Drone moved to (2, 5) Accumulated
Reward: -190
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
```



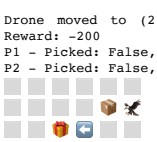
```
Drone moved to (2, 4) Accumulated
Reward: -190
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
```



```
Drone moved to (2, 3) Accumulated
Reward: -200
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
```

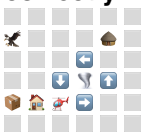


```
Drone moved to (2, 4) Accumulated
Reward: -200
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
```



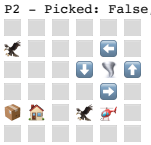
```
Drone moved to (2, 5) Accumulated
Reward: -200
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
```

Random Drone movement in stochastic env for 10+ steps to show that the environment logic is defined correctly.



Attempted dropoff failed (no package carried)
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Drone moved to (3, 4) Accumulated Reward: -50
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False

Drone moved to (4, 4) Accumulated Reward: -50
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False



Attempted dropoff failed (no package carried)
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Drone moved to (5, 4) Accumulated Reward: -50
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False

Attempted dropoff failed (no package carried)
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Drone moved to (4, 4) Accumulated Reward: -150
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False

Drone moved to (3, 4) Accumulated Reward: -200
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False

Attempted dropoff failed (no package carried)
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Drone moved to (4, 4) Accumulated Reward: -210
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False

Drone moved to (4, 3) Accumulated Reward: -210
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False

Drone moved to (5, 3) Accumulated Reward: -210

P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False

Attempted dropoff failed (no package carried)
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Drone moved to (4, 3) Accumulated Reward: -210
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False

Attempted dropoff failed (no package carried)
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Attempted dropoff failed (no package carried)
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Picked up package 1
P1 - Picked: True, Delivered: False
P2 - Picked: False, Delivered: False
Drone moved to (4, 4) Accumulated Reward: -210
P1 - Picked: True, Delivered: False
P2 - Picked: False, Delivered: False

Dropped package 1 incorrectly
Attempted dropoff failed (no package carried)
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Drone moved to (4, 3) Accumulated Reward: -210
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False

Drone moved to (5, 3) Accumulated Reward: -210
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False

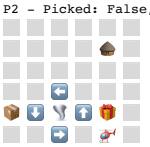
Drone moved to (5, 4) Accumulated Reward: -210
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False

Drone moved to (5, 3) Accumulated Reward: -210
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False

Drone moved to (5, 4) Accumulated Reward: -210
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False

Drone moved to (4, 4) Accumulated Reward: -210
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False

Drone moved to (4, 5) Accumulated Reward: -210
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False



Drone moved to (3, 5) Accumulated Reward: -220
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False

Drone moved to (2, 5) Accumulated Reward: -230
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False

Attempted dropoff failed (no package carried)
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Attempted pickup failed
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Attempted dropoff failed (no package carried)
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Drone moved to (3, 5) Accumulated Reward: -240
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False

Drone moved to (4, 5) Accumulated Reward: -240
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False

Drone moved to (5, 5) Accumulated Reward: -240
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False

Attempted dropoff failed (no package carried)
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Attempted pickup failed
P1 - Picked: False, Delivered: False

P2 - Picked: False, Delivered: False
Attempted dropoff failed (no package carried)
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Attempted dropoff failed (no package carried)
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Drone moved to (5, 4) Accumulated Reward: -240
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False

Attempted dropoff failed (no package carried)
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Attempted dropoff failed (no package carried)
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Attempted dropoff failed (no package carried)
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Attempted dropoff failed (no package carried)
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Drone moved to (4, 4) Accumulated Reward: -240
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False

Drone moved to (4, 3) Accumulated Reward: -240
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False

Drone moved to (4, 4) Accumulated Reward: -240
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False

Drone moved to (4, 3) Accumulated Reward: -240
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False

Attempted dropoff failed (no package carried)
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False
Drone moved to (5, 3) Accumulated Reward: -240
P1 - Picked: False, Delivered: False
P2 - Picked: False, Delivered: False