

# 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/environments/toy_text/)
- [https://gymnasium.farama.org/tutorials/gymnasium\\_basics/environment\\_creation/#sphx-glr-tutorials-gymnasium-basics-environment-creation-py](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

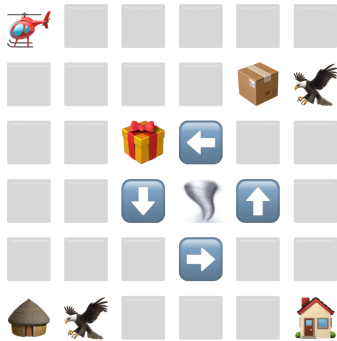
**Rewards System:**  
The drone earns or loses points based on its actions and performance.

| 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, <b>not yet implemented</b> , 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. 🚁✅



|   |   |   |   |   |
|---|---|---|---|---|
|   |   |  |   |   |
|   |   |   |  |  |
|   |  |  |   |   |
|   |   |  |  |  |
|   |   |   |  |   |
|  |  |   |   |  |

|   |   |    |    |    |   |   |  |
|---|---|----|----|----|---|---|--|
|   |   |    |    |    |   |   |  |
|   |   |    |    | 🚁  |   |   |  |
|   |   |    |    |    | 📦 | 🦇 |  |
|   |   | 🎁  | ⬅️ |    |   |   |  |
|   |   | ⬇️ | 👁️ | ⬆️ |   |   |  |
|   |   |    | ➡️ |    |   |   |  |
| 🏠 | 🦇 |    |    |    |   | 🏡 |  |

|   |   |   |   |   |   |
|---|---|---|---|---|---|
|   |   |  |   |   |   |
|   |   |   |   |  |  |
|   |   |  |  |   |   |
|   |   |   |   |   |   |
|   |   |  |  |  |   |
|   |   |   |   |   |   |
|   |   |   |  |   |   |
|  |  |   |   |   |  |

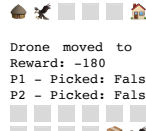
[illegible]

A 10x10 grid world environment. The grid contains several objects and obstacles: a red helicopter in the top row, a brown box in the top row, a black eagle in the top row, a red gift box in the second row, a blue left arrow in the second row, a blue down arrow in the third row, a blue up arrow in the third row, a blue right arrow in the third row, a brown house in the bottom row, a black eagle in the bottom row, and a red house in the bottom row. The grid is otherwise empty.

A 10x10 grid world environment. The grid contains several objects and obstacles: a brown box in the top right, a black eagle in the top right, a red gift box in the middle left, a red helicopter in the middle left, a blue square with a white down arrow in the middle left, a grey elephant in the middle left, a blue square with a white up arrow in the middle right, a blue square with a white right arrow in the bottom middle, a brown hut in the bottom left, a black eagle in the bottom left, and a red house in the bottom right.

A 5x5 grid world environment. The grid contains several objects and obstacles: a red shield with a yellow star in the second row, second column; a red and white striped rocket in the third row, second column; a blue square with a white left arrow in the third row, third column; a blue square with a white up arrow in the fourth row, third column; a blue square with a white right arrow in the fifth row, third column; a brown cardboard box in the first row, fourth column; a black dragon in the first row, fifth column; a brown house in the fifth row, first column; a black dragon in the fifth row, second column; and a red house in the fifth row, fifth column.

A 10x10 grid world environment. The grid contains several objects and obstacles: a brown box in the top right, a black horse in the top right, a red gift box in the middle left, a blue left arrow in the middle left, a blue down arrow in the middle left, a blue up arrow in the middle right, a blue right arrow in the bottom middle, a red and blue robot in the bottom middle, a brown house in the bottom left, a black horse in the bottom left, and a red and blue house in the bottom right.



Attempted dropoff  
carried)  
P1 - Picked: False  
P2 - Picked: False  
Attempted dropoff  
carried)

A 10x10 grid world environment. The grid contains several objects and obstacles: a brown house in the bottom-left corner (row 9, column 1); a black dragon in the bottom-left area (row 9, column 2); a red gift box in the middle-left area (row 5, column 4); a blue square button with a white left arrow in the middle (row 5, column 5); a blue square button with a white down arrow in the middle-left (row 6, column 4); a grey snake in the middle (row 6, column 5); a blue square button with a white up arrow in the middle-right (row 6, column 6); a blue square button with a white right arrow in the bottom-middle (row 7, column 5); a brown cardboard box in the top-right area (row 2, column 8); a black dragon in the top-right area (row 2, column 9); a red helicopter in the bottom-right area (row 9, column 8); and a red house in the bottom-right corner (row 9, column 9).

A 10x10 grid world environment. The grid contains several objects and obstacles: a brown house in the bottom-left corner (row 9, column 1); a black dragon in the bottom-left area (row 9, column 3); a red helicopter in the bottom-middle (row 8, column 6); a blue house in the bottom-right corner (row 9, column 10); a blue square with a white arrow pointing up in the middle-right (row 6, column 8); a blue square with a white arrow pointing down in the middle-left (row 5, column 4); a blue square with a white arrow pointing left in the middle-left (row 5, column 5); a red gift box in the middle-left (row 5, column 6); a grey smokestack in the middle-right (row 6, column 6); a brown cardboard box in the top-right area (row 2, column 8); and a black dragon in the top-right corner (row 2, column 10). All other cells in the grid are empty.

A 10x10 grid world environment. The grid contains several objects and obstacles: a brown house in the bottom-left corner (row 9, column 1); a black dragon in the bottom-left area (row 9, column 3); a red helicopter in the bottom-center (row 9, column 6); a red house in the bottom-right corner (row 9, column 10); a brown box in the top-right area (row 2, column 8); a red gift box in the center (row 5, column 4); a blue square with a white left arrow (row 5, column 5); a blue square with a white down arrow (row 6, column 4); a grey snake in the center (row 6, column 5); a blue square with a white up arrow (row 6, column 6); and a blue square with a white right arrow (row 7, column 5). All other cells in the grid are empty.

A 5x5 grid world environment. The grid contains the following elements:

- Obstacles (Gray Squares):** Located at (0,0), (0,1), (0,2), (0,3), (0,4), (1,0), (1,1), (1,2), (1,3), (1,4), (2,0), (2,1), (2,2), (2,3), (2,4), (3,0), (3,1), (3,2), (3,3), (3,4), (4,0), (4,1), (4,2), (4,3), (4,4).
- Start (Green Circle):** Located at (1,1).
- Goal (Red Star):** Located at (4,4).
- Items:**
  - Red gift box at (1,2).
  - Blue left arrow at (1,3).
  - Blue down arrow at (1,4).
  - Blue up arrow at (2,3).
  - Blue right arrow at (2,4).
  - Gray key at (2,2).
  - Brown box at (3,2).
  - Black dragon at (3,3).
  - Red helicopter at (3,4).
  - Yellow house at (4,3).
- Other:** A small brown house is located at (0,4).

A 10x10 grid world environment. The grid contains several objects and obstacles: a brown box in the top right, a black dragon in the top right, a red gift box in the middle left, a blue left arrow in the middle left, a blue down arrow in the middle left, a grey snake in the middle left, a blue up arrow in the middle left, a red helicopter in the bottom left, a blue right arrow in the bottom left, a brown house in the bottom left, a black dragon in the bottom left, and a red house in the bottom right.

A 10x10 grid world environment. The grid contains several objects and obstacles: a brown box in the top right, a black dragon in the top right, a red gift box in the middle left, a blue left arrow in the middle left, a blue down arrow in the middle left, a blue up arrow in the middle right, a blue right arrow in the bottom middle, a small brown house in the bottom left, a black dragon in the bottom left, a red helicopter in the bottom middle, and a small red house in the bottom right.

A 10x10 grid world environment. The grid contains several objects and obstacles: a brown house in the bottom-left corner (row 9, column 1); a black dragon in the bottom-left corner (row 9, column 2); a red gift box in the middle-left area (row 5, column 4); a red helicopter in the middle-left area (row 6, column 4); a blue square with a white left arrow in the middle-right area (row 5, column 6); a blue square with a white right arrow in the bottom-middle area (row 8, column 6); a blue square with a white up arrow in the middle-right area (row 6, column 7); a brown cube in the top-right area (row 2, column 8); a black dragon in the top-right area (row 2, column 9); and a red house in the bottom-right corner (row 9, column 10).

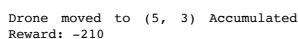
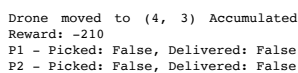
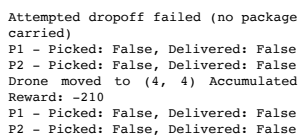
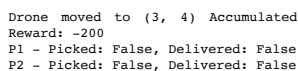
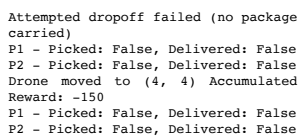
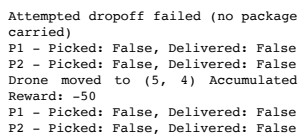
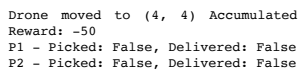
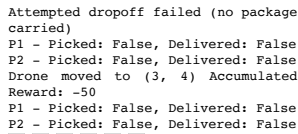
A 10x10 grid world environment. The grid contains several objects and obstacles: a brown house in the bottom-left corner (row 9, column 1); a black eagle in the bottom-left area (row 9, column 2); a red helicopter in the bottom-center (row 9, column 4); a blue house in the bottom-right corner (row 9, column 10); a brown box in the top-right area (row 2, column 8); a red gift box in the center (row 5, column 4); a blue arrow pointing left in the center (row 5, column 5); a blue arrow pointing down in the center (row 6, column 4); a blue arrow pointing up in the center (row 6, column 6); a blue arrow pointing right in the center (row 7, column 5); and a grey elephant in the center (row 6, column 5). All other cells in the grid are empty.

A 5x5 grid world environment. The grid contains the following objects:

- Top-left: A black eagle.
- Top-right: A small brown house.
- Middle-left: A brown cardboard box.
- Middle: A red house.
- Middle-right: A red helicopter.
- Bottom-left: A blue arrow pointing down.
- Bottom-middle: A grey tornado.
- Bottom-right: A blue arrow pointing up.

Navigation arrows are located in the center of the grid:

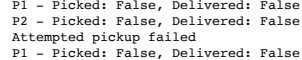
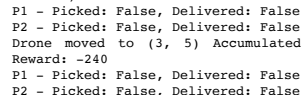
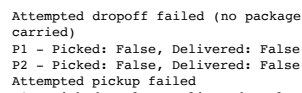
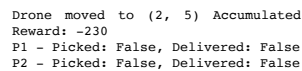
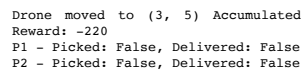
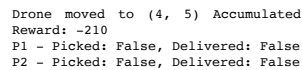
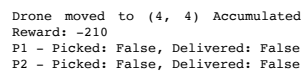
- Top-middle: A blue arrow pointing left.
- Middle-right: A blue arrow pointing right.



```
P1 - Picked: False
P2 - Picked: False
Picked up package
P1 - Picked: True
P2 - Picked: False
Drone moved to
Reward: -210
P1 - Picked: True
```

|   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

|   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |



```
P1 - Picked: False
P2 - Picked: False
Drone moved to
Reward: -240
P1 - Picked: False
P2 - Picked: False
```

```
P1 - Picked: False
P2 - Picked: False
Attempted dropoff:
  carried)
P1 - Picked: False
P2 - Picked: False
Attempted dropoff:
  carried)
```

```










Attempted dropoff
carried)
P1 - Picked: False
P2 - Picked: False
Drone moved to
Reward: -240
P1 - Picked: False
P2 - Picked: False

```

A 6x6 grid world environment. The grid contains several objects and obstacles: a brown box in the bottom-left corner (row 5, column 1); a small house in the bottom-left corner (row 5, column 2); a blue square with a white arrow pointing right in the bottom-left corner (row 5, column 3); a blue square with a white arrow pointing down in the middle-left (row 3, column 2); a blue square with a white arrow pointing left in the middle-left (row 3, column 3); a blue square with a white arrow pointing up in the middle-right (row 3, column 5); a blue square with a white arrow pointing right in the middle-right (row 3, column 6); a brown house in the top-right (row 1, column 5); a red gift box in the bottom-right (row 5, column 6); and a black bird in the top-right corner (row 1, column 6). The rest of the grid cells are empty.

Drone moved to  
Reward: -240  
P1 - Picked: False  
P2 - Picked: False

P1 - Picked: False  
P2 - Picked: False

|   |   |   |  |   |   |
|---|---|---|--|---|---|
|   |   |   |  |   |  |
|   |   |   |  |  |   |
|   |  |   |  |   |   |
|  |  |  |  |   |  |
|  |  |   |  |  |   |

|   |   |   |   |   |   |
|---|---|---|---|---|---|
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |