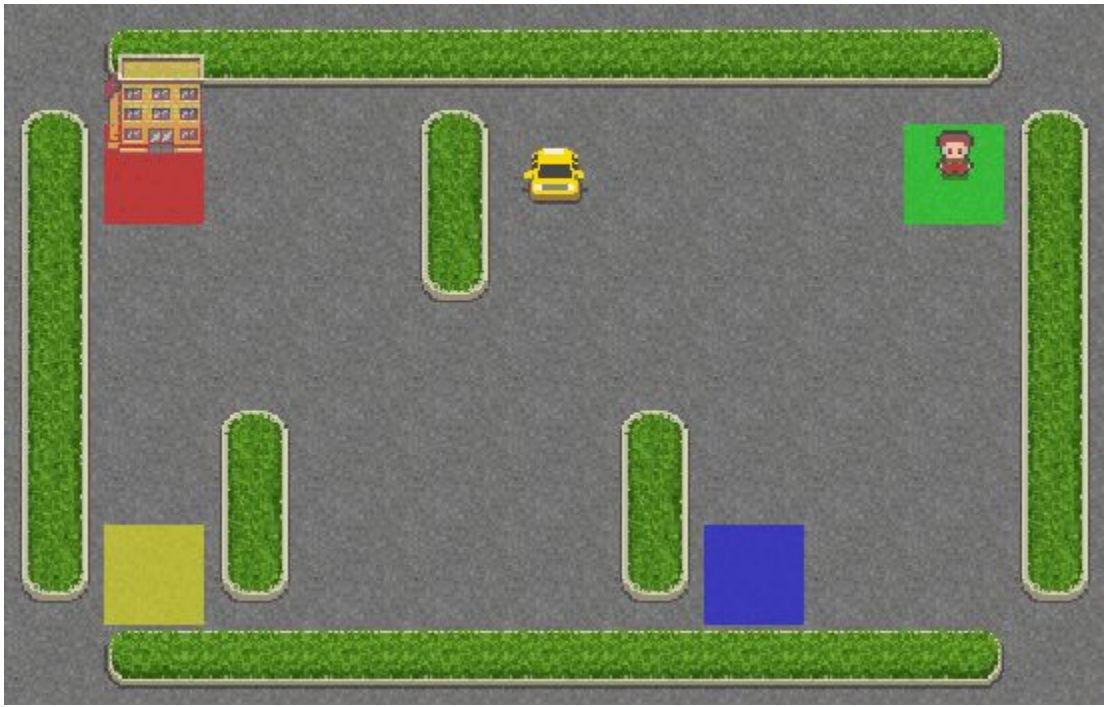
AIL 722
Assignment 2

Vaibhav Bihani & Sanket Gandhi

Overview

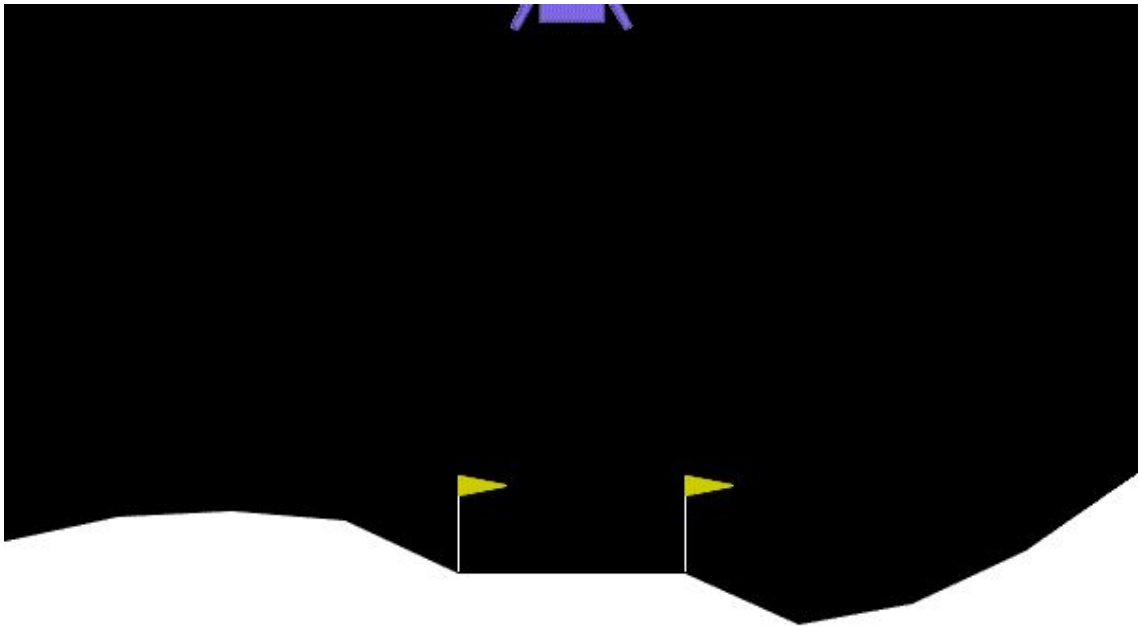
- Three Parts:
  - Model based methods: Value Iteration, Policy Iteration [20]
  - Model free methods: SARSA, Q-Learning [40]
  - Large and continuous states space methods : Deep Q-Networks [20+20]
- Environments:
  - Taxi-v3
  - LunarLander-v2
  - TreasureHunt
- Due Date : 17th October 11:55 PM
- A Viva will be conducted

Environments: Taxi-v3



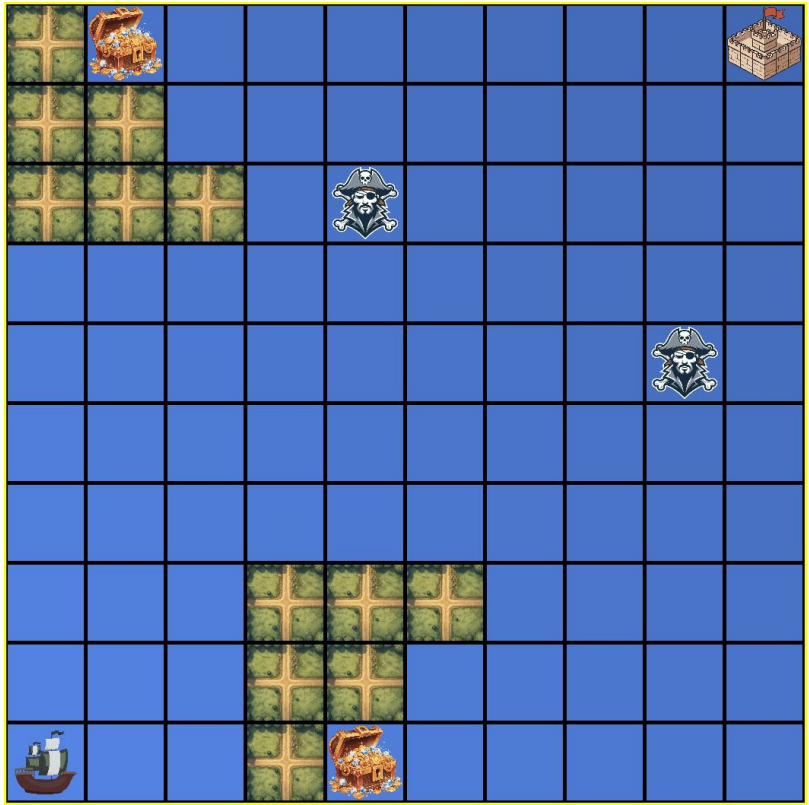
- **Goal:** Navigate a taxi to pick up and drop off passenger
- **State space:** 500 discrete states
- **Action space:** 6 actions (move in 4 directions, pick up, drop off).
- **Reward:** +20 for successful drop-off, -1 per step, and -10 for illegal actions.

Environments: LunarLander-v2



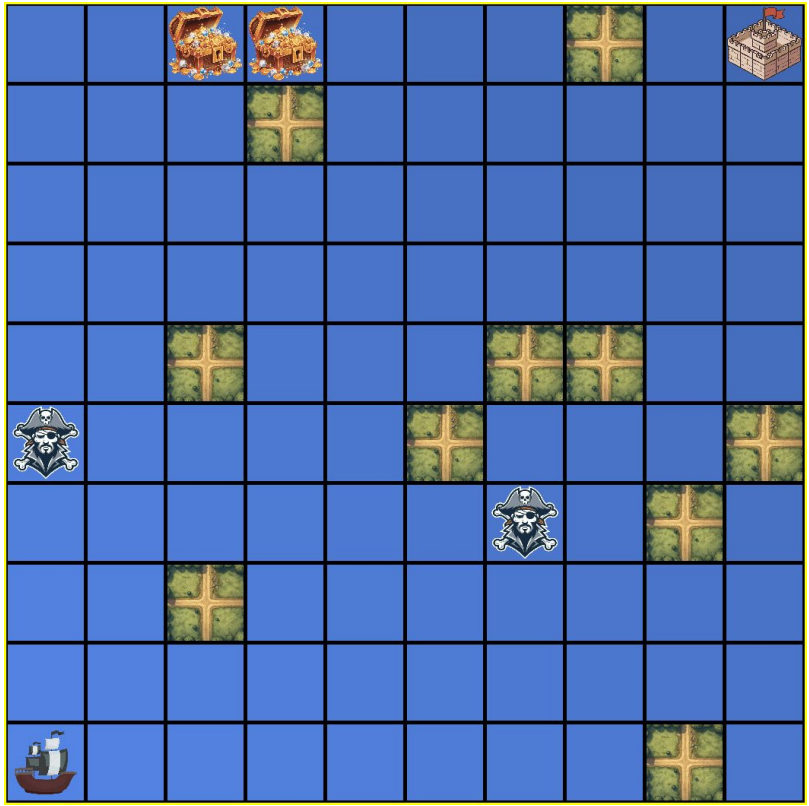
- **Goal**: Safely land a lunar module on a designated landing pad.
- **State space**: 8 continuous variables (module's position, velocity, angle, etc.).
- **Action space**: 4 discrete actions (do nothing, fire left, right, or main engine).
- **Reward**: Based on distance to landing pad, speed, tilt, engine use etc. , with +100 for landing, -100 for crashing, and 200+ points to solve an episode.

Environments: TreasureHunt



- **Goal**: Take boat to the fort while avoiding pirates and picking up treasure
- **State space**: 100*4 states discrete states
- **Action space**: 4 discrete actions (move right, left, top and bottom).
- **Reward**: +2 for treasure -1 for pirate and +0.25 for fort. -0.1 for living

Environments: TreasureHunt-v2



- **Goal**: Take boat to the fort while avoiding pirates and picking up treasure
- **State space**: of the order 10^10
- **Action space**: 4 discrete actions (move right, left, top and bottom).
- **Reward**: +2 for treasure -1 for pirate and +0.25 for fort. -0.1 for living

Algorithms and Environments

Environments:

Environments	Transitions	State Space	Action Space	Algorithms
Taxi-v3	Unknown	Discrete	Discrete	
TreasureHunt-v1	Known	Discrete	Discrete	
TreasureHunt-v1	Unknown	Discrete	Discrete	
Lunar-Lander-v2	Unknown	Continuous	Discrete	
TreasureHunt-v2	Known	Discrete but Large	Discrete	

Algorithms and Environments

Environments:

Environments	Transitions	State Space	Action Space	Algorithms
Taxi-v3	Unknown	Discrete	Discrete	SARSA, Q-learning
TreasureHunt-v1	Known	Discrete	Discrete	
TreasureHunt-v1	Unknown	Discrete	Discrete	
Lunar-Lander-v2	Unknown	Continuous	Discrete	
TreasureHunt-v2	Known	Discrete but Large	Discrete	

Algorithms and Environments

Environments:

Environments	Transitions	State Space	Action Space	Algorithms
Taxi-v3	Unknown	Discrete	Discrete	SARSA, Q-learning
TreasureHunt-v1	Known	Discrete	Discrete	VI, PI
TreasureHunt-v1	Unknown	Discrete	Discrete	
Lunar-Lander-v2	Unknown	Continuous	Discrete	
TreasureHunt-v2	Known	Discrete but Large	Discrete	

Algorithms and Environments

Environments:

Environments	Transitions	State Space	Action Space	Algorithms
Taxi-v3	Unknown	Discrete	Discrete	SARSA, Q-learning
TreasureHunt-v1	Known	Discrete	Discrete	VI, PI
TreasureHunt-v1	Unknown	Discrete	Discrete	SARSA, Q-learning
Lunar-Lander-v2	Unknown	Continuous	Discrete	
TreasureHunt-v2	Known	Discrete but Large	Discrete	

Algorithms and Environments

Environments:

Environments	Transitions	State Space	Action Space	Algorithms
Taxi-v3	Unknown	Discrete	Discrete	SARSA, Q-learning
TreasureHunt-v1	Known	Discrete	Discrete	VI, PI
TreasureHunt-v1	Unknown	Discrete	Discrete	SARSA, Q-learning
Lunar-Lander-v2	Unknown	Continuous	Discrete	DQN
TreasureHunt-v2	Known	Discrete but Large	Discrete	

Algorithms and Environments

Environments:

Environments	Transitions	State Space	Action Space	Algorithms
Taxi-v3	Unknown	Discrete	Discrete	SARSA, Q-learning
TreasureHunt-v1	Known	Discrete	Discrete	VI, PI
TreasureHunt-v1	Unknown	Discrete	Discrete	SARSA, Q-learning
Lunar-Lander-v2	Unknown	Continuous	Discrete	DQN
TreasureHunt-v2	Known	Discrete but Large	Discrete	DQN

HPC IIT Delhi

https://supercomputing.iitd.ac.in/

- Windows/Mac doesn't support all of the gymnasium functionalities. Hence recommended to use HPC for this assignment

- Setup Instructions are given in assignment document
- Requires setting up a user account
- Computational resources are charged:

Standard Queue

CPU	₹ 0.10/hour/core
GPU	₹ 1.00/hour/card

High Queue:

CPU	₹ 0.50/hour/core
GPU	₹ 5.00/hour/card

Each student has been allotted Rs. **165.00** through **'ail722.{kerberos}.course**' project . You need to **utilise it for both A2 and A3**.

DQN Implementation

- For DQN Implementation , Q-Network Architecture is predefined and provided along with the code
- You need to write the algorithm and training code
- Evaluation and Visualisation scripts are provided as well.
- Ensure saving checkpoints
- Training time ~30 mins(1 CPU) for LunarLander-v2, and ~5 hr for TreasureHunt

Code Template

Thank You