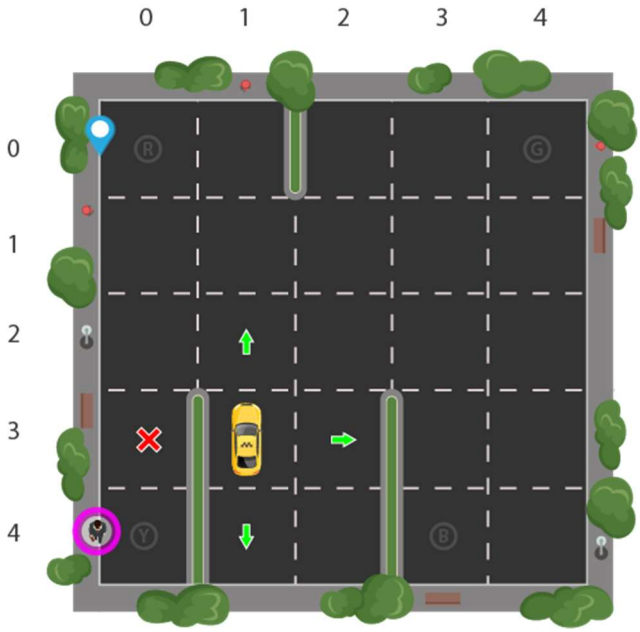
Student Name : Rakyan Adhikara

Student ID : 219548135

SIT215 Project (Report)

1. The Taxi Problem:
    - Definition:
        - The main task for the taxi problem is to pick up passengers at one location and drop them off on another location. Here are a few things that needs to be consider about the taxi problem:
            - When the taxi arrived at the destination, drop the passenger off.
            - Take as small amount of time as possible to save the passenger's time to drop off.
            - Take care of passenger's safety and traffic rules
        - The goal is to pick up a passenger at one of the 4 possible locations and to drop him off in another.
    - Mathematical Problem:
        - Reinforced Learning:
            - According to Kansal and Martin, Reinforcement Learning is the science of making optimal decisions using experiences. (2018)
        - Kansal and Martin (2018) also states that there are three key elements that need to be considered on this problem:
            - Rewards
                - Since it is highly desired, the agent should earn a high positive reward for each successful passenger drop-off.
                - In order to prevent any bad customer service, the agent should get a penalty if it ever tries to drop the passenger off in wrong directions.
                - The agent should get a small amount of negative reward for not making it to the destination after every time-step. The reason being is it is tolerable for our agent to reach the destination late instead of making wrong moves while trying to reach to the destination as fast as possible.
            - States

- The taxi environment has 5 × 5 × 5 × 4 = 500 total possible states.
    i. For instance, the training area for the taxi to take passengers and drop them off in a parking lot to four different locations (R, G, Y, B):

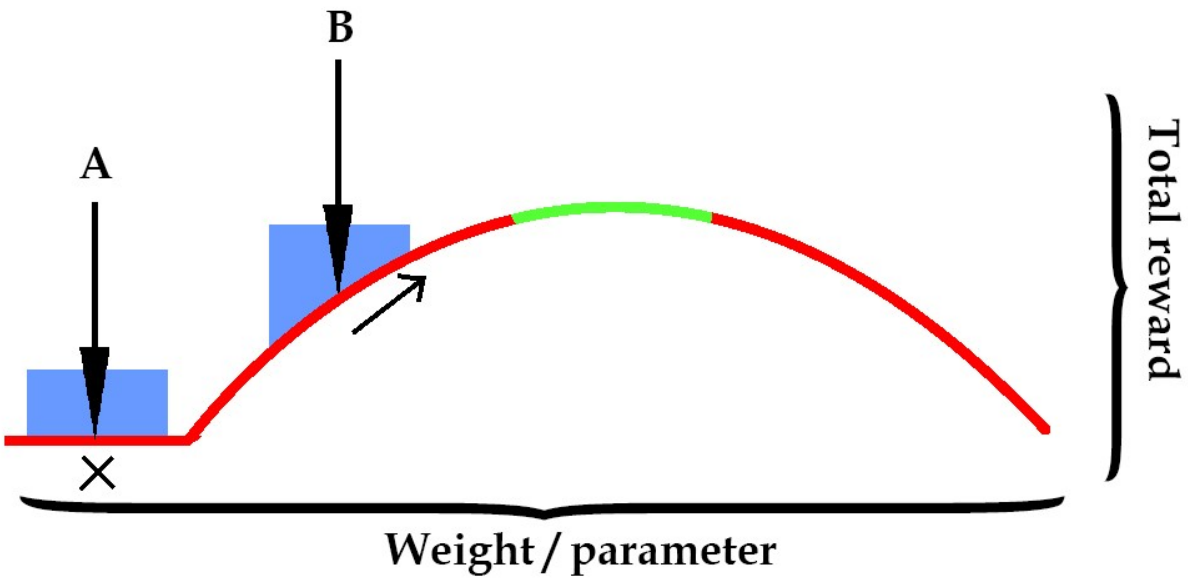


    ii. Assume that the taxi is the only vehicle in this parking lot. Split the parking lot into a 5x5 grid, which gives us 25 possible taxi locations.
    iii. Notice that there are four pinpoints or locations that the taxi can pick up and drop a passenger off: R, G, Y, B
    iv. When also account for one additional passenger state of being inside the taxi, the taxi can take all combinations of passenger locations and their destination locations to come to a total number of states for our taxi environment, which in total, there are four destinations and five (4 + 1) passenger locations.
    v. Thus, the taxi environment has 5 × 5 × 5 × 4 = 500 total possible states.

- Actions
    - Every time the agent encounters one of the 500 states, it takes an action.
    - There are six possible actions:
        i. South
        ii. North
        iii. East
        iv. West
        v. PickUp
        vi. DropOff
    - The actions in this case are to move in a direction, to decide to pick-up a passenger and to drop them off on the destination.

- Solution:
    - Q-Learning:
        - Q-Learning is a reinforcement learning (RL) algorithm which seeks to find the best action the agent should take given the current state. The main goal is to identify a policy that maximises the expected overall reward.
    - According to Kansai and Martin (2018), there are 7 steps:
        - Initialize the Q-table's values to zeros.
        - Start exploring actions: For each state, select any one among all possible actions for the current state (S).
        - Travel to the next state (S') as a result of that action (a).
        - For all possible actions from the state (S') select the one with the highest Q-value.
        - Update Q-table values using the equation.
        - Set the next state as the current state.
        - If goal state is reached, then end and repeat the process.

2. The Cart-Pole Problem:
    - Definition:
        - Also known as “Inverted pendulum”
        - Pendulum with a center of gravity above its pivot point.
        - According to Surma (2018), the main goal is to keep the cartpole as balanced as possible by applying any appropriate forces to a pivot point.
    - Mathematical Problem:
        - The rewards, states and actions are:
            - Rewards:
                - A reward of +1 is provided for every timestep that the pole remains upright.
            - States:
                - There are +1 states every time the weight got increased
            - Actions:
                - There are two possible actions:
                    - Left
                    - Right
    - Solution:
        - Hill-Climbing
            - Main Idea: Instead of Trial and Error and see if any combination works, gradually improve the weights.



- According to Frans (2016), based on the picture above,
    - An arrow represents the value that weights are initialized at. The blue region is noise that is added at every iteration.
    - If the weight starts at arrow B, then hill-climbing can try other weights near arrow B until it finds one that improves the reward. This can continue until the weight is in the green zone.
    - However, arrow A has a problem. If the weight is initialized at a location where changing it slightly gives no improvement at all, the hill climbing is stuck and won't find a good answer.

3. In Comparison…

Problem-type	Definition	Mathematical Problem	Q-Learning Solution
The Taxi Problem	The main task for the taxi problem is to pick up passengers at one location and drop them off on another location. The goal is to pick up a passenger at one of the 4 possible locations and to drop him off in another.	Rewards: - Receive a high positive reward for a successful drop-off. - Penalized every time the taxi tries to drop off a passenger in wrong locations. - Get a small amount of negative reward for not making it to the destination after every time-step. States: There are 500 possible states. Actions: There are six possible actions: - South - North - East - West - Pickup - Dropoff	- Initialize the Q-table's values to zeros. - Start exploring actions: For each state, select any one among all possible actions for the current state (S). - Travel to the next state (S') as a result of that action (a). - For all possible actions from the state (S') select the one with the highest Q-value. - Update Q-table values using the equation. - Set the next state as the current state. - If goal state is reached, then end and repeat the process.

The Cart-Pole Problem	Pendulum with a center of gravity above its pivot point. The main goal is to keep the cartpole as balanced as possible by applying any appropriate forces to a pivot point	Rewards: A reward of +1 is provided for every timestep that the pole remains upright. States: There are +1 states every time the weight got increased Actions: There are two possible actions: - Left - Right	Hill-Climbing: - Main idea: Instead of Trial and Error and see if any combination works, gradually improve the weights.

GitHub Repository: https://github.com/rkynaa/ReinforcedLearning

References:

- Kansal, S & Martin, B 2018, *Reinforcement Q-Learning from Scratch in Python with OpenAI Gym*, LearnDataSci, Retrieved 29th September 2020, < https://www.learndatasci.com/tutorials/reinforcement-q-learning-scratch-python-openai-gym/>
- Sukma, G 2018, *Cartpole - Introduction to Reinforcement Learning (DQN - Deep Q-Learning)*, Towards Data Science, Retrieved 29th September 2020, < https://towardsdatascience.com/cartpole-introduction-to-reinforcement-learning-ed0eb5b58288>
- Frans, K 2016, *Simple reinforcement learning methods to learn CartPole*, Kevin Frans, Retrieved 29th September 2020 < http://kvfrans.com/simple-algoritms-for-solving-cartpole/>