**Vivekanand Education Society’s Institute of Technology**

**Department of AI&DS Engineering**



**Subject: Reinforcement Learning**

**Class: D16AD**

| ROLL NO: **30** | NAME: [**SUHANEE KANDALKAR**](mailto:2021.suhanee.kandalkar@ves.ac.in) | | |
| --- | --- | --- | --- |
| EXPERIMENT NO:**1** | TITLE:Implement a simple grid-world environment and train an agent using basic Q learning using python. | | |
| DOP: |  | DOS: **25/01/25** |  |
| GRADES: | LOs MAPPED: | | SIGNATURE: |

## Aim**:**

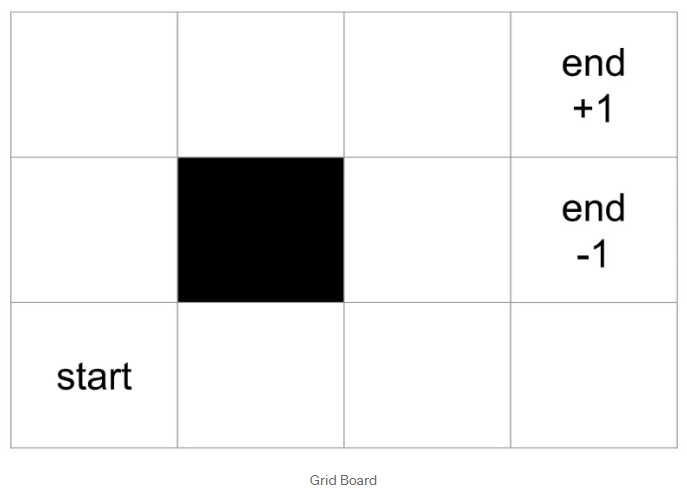
Implement a simple grid-world environment and train an agent using basic Q learning using python.

## Theory**:**

Reinforcement Learning (RL) is a branch of machine learning where agents learn to make decisions by interacting with an environment to maximize cumulative rewards.  
 This experiment focuses on implementing a **simple grid-world environment** and training an agent using the **Q-Learning algorithm.**[**Reinforcement Learning — Implement Grid World**](https://towardsdatascience.com/reinforcement-learning-implement-grid-world-from-scratch-c5963765ebff)

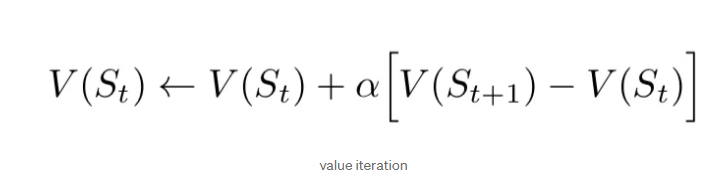
### Simple Grid-World Environment

* A grid-world is a small, two-dimensional grid where each square (or cell) represents a possible position for an agent (like a robot or character).
* The agent can move in four directions: up, down, left, or right.
* Some cells have rewards (like reaching a goal), while others may have penalties (like hitting an obstacle).
* The goal is for the agent to navigate the grid and maximize the rewards it collects while minimizing penalties.
* Think of it like a board game where the agent learns the best path to win over time.



a **policy** is a mapping from state to action

**Value iteration**, just as its name, update its value(estimated reward) at each iteration(end of game)**.**

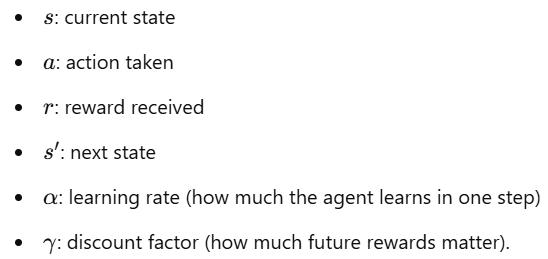


Here is one last thing we need to talk about. Once our agent finds a path to get reward +1, should it sticks to it and forever follows that path (exploitation)   
Or  
 should it gives other path a chance(exploration) and expects a shorter path?

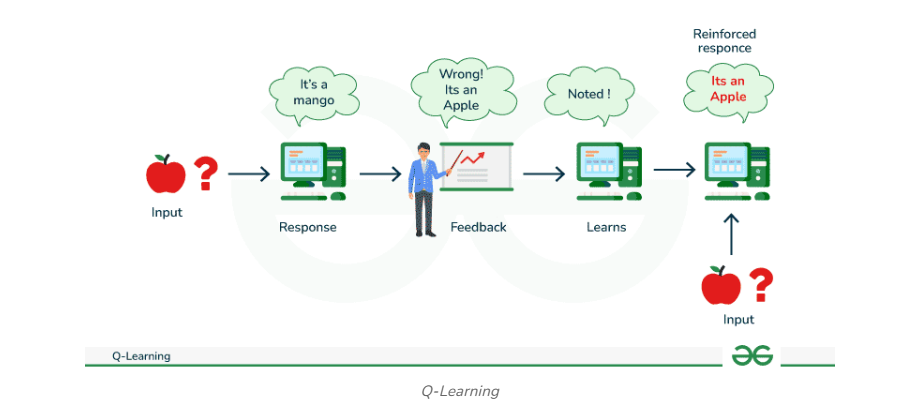
**Deterministic**. This means that when the agent selects an action (e.g., "move right"), it always moves exactly in that direction without randomness.

### Basic Q-Learning

* Q-Learning is a simple reinforcement learning algorithm that teaches an agent how to act in an environment.
* It uses a Q-table, where each entry represents the "quality" (Q-value) of taking a specific action in a given state.
* The agent updates the Q-values based on rewards it receives, using this formula:

****

Over time, the agent learns to pick actions that lead to the highest rewards.



[Q-Learning Geeks For Geeks](https://www.geeksforgeeks.org/q-learning-in-python/?utm_source=chatgpt.com)

[**Q-Learning Algorithm: From Explanation to Implementation**](https://towardsdatascience.com/q-learning-algorithm-from-explanation-to-implementation-cdbeda2ea187)

# Code:

[RL\_exp1\_30\_Code\_Implementation](https://docs.google.com/document/d/1ayBGZfZCXDWdmWFIsZ6ATH_aeW59coPXXiBSg2k2DUk/edit?usp=sharing)

# Conclusion :

The agent learns to navigate a grid-world environment by iteratively updating its Q-values based on the rewards received for different actions. Through this process, the agent gradually improves its policy, balancing exploration and exploitation to reach the goal.