# Introduction to Artificial Intelligence

**Assignment 4**

# Reinforcement Learning

## Lab Report

## Group 49

### Group Members


| Name | Index No. |
|---|---|
| AMRITHSHAGAR J. | 200037X |
| PAHASARA G.G.V. | 200440C |
| THARINDA N.H.D. | 200638P |
| YAPA Y.M.C.C.L.C. | 200742E |

# Introduction

According to [1], “Reinforcement Learning(RL) is a type of machine learning technique that enables an agent to learn in an interactive environment by trial and error using feedback from its own actions and experiences”. RL can be subdivided into Model-based and Model-free learning. Model-free learning can be divided into active and passive reinforcement learning. Q-learning is a technique used in active, model-free RL. As can be understood from the word, model-free learning is when an agent learns without a transition and reward model of the environment it is acting on. Active RL is when the agent is not given a fixed policy and is free to explore the environment.

# Experiment

In this report, the policy convergence of Q-Learning is compared with different learning and discount rates with the Taxi environment of OpenAI’s Gym. The code and tutorial given in [2] are used for the simulation with various learning and discount rates. Our task is to comment on the influence the above parameters have on how fast Q-learning can converge.

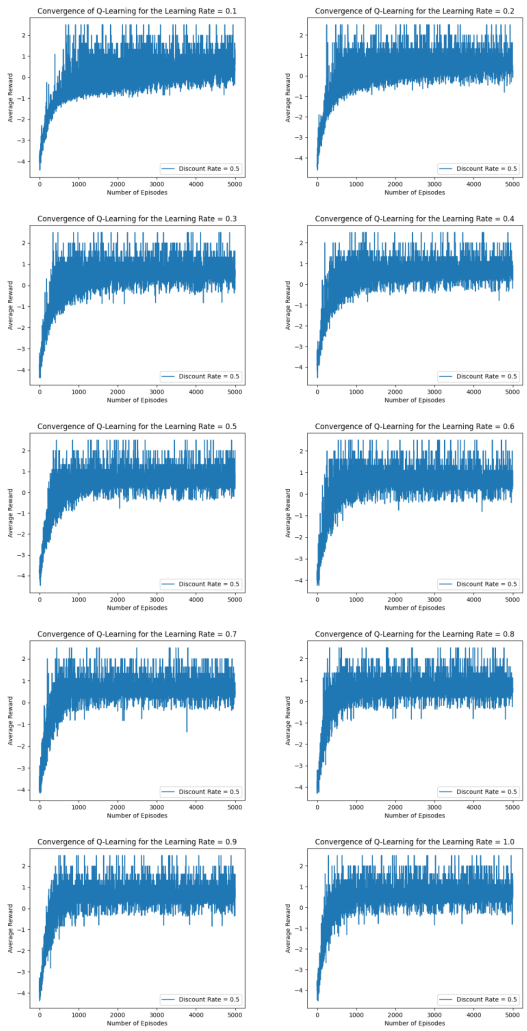
The experiment was conducted in two stages. In the first stage, the discount rate was kept fixed at 0.5 while the learning rate was increased from 0.1 to 1.0 by steps of 0.1. In the second stage, the learning rate was kept fixed at 0.5 while the discount rate was increased from 0.1 to 1.0 by steps of 0.1. The code for this experiment is given here.

# Observations

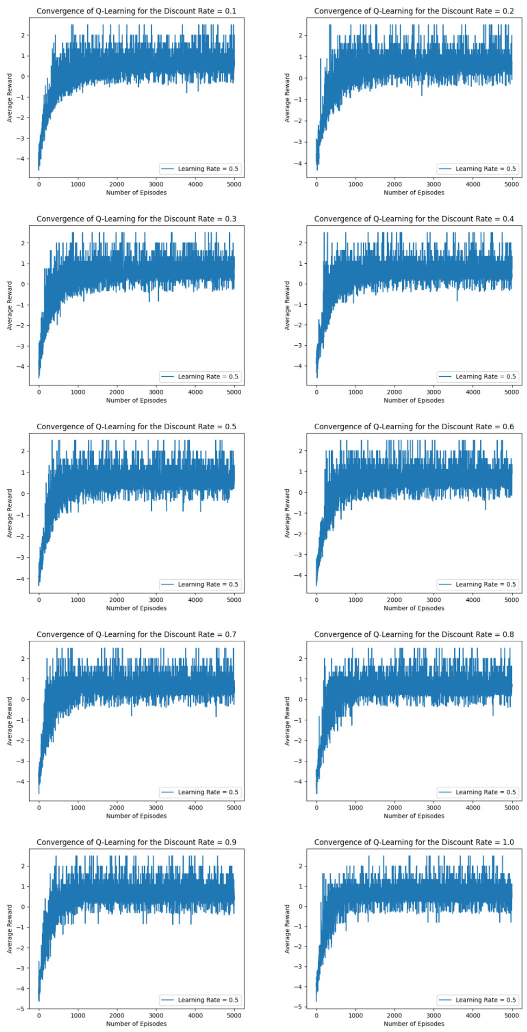
Plotted diagrams have the number of episodes in the x-axis and the average reward per each episode in the y-axis.

When the average reward becomes a constant, the agent is not getting any new learning data, which implies the convergence of Q-learning. More observations are given here.

**Graphs obtained at the first stage. (fixed discount rate and variable learning rate)**




**Graphs obtained at the second stage. (fixed discount rate and variable learning rate)**



As per the first set of graphs, The No. of episodes taken to the Convergence of the Q-learning algorithm decreases with the increased learning rate.
As per the second set of graphs, The No. of episodes taken to the convergence of the Q-learning algorithm decreases with the discount rate increase.

## Conclusion

In low learning rates, the agent learns new data slowly. As in graphs, when the learning rate is high, the agent takes longer to converge than when the learning rate is higher. When the learning rate is low, Q-learning convergence gets low, which implies that for higher-valued learning rates, Q-learning converges faster, considering the lower values.

The agent learns new data at high discount rates more quickly than when the discount rate is lower. As in the graphs, the rate of convergence of Q-Learning gets low when the discount rate is getting low. Q-learning converges more quickly when the discount rate is high and low when it is low.

When comparing the graphs of learning rates of 0.9 and 1.0, the average reward oscillates more frequently in the graph with a learning rate of 1.0 than in the chart with a learning rate of 0.9. Therefore, If the learning rate is too high, the agent may overreact to new information and oscillate in its decision-making, leading to slow convergence or divergence. On the other hand, if the learning rate is too low, the agent may take too long to incorporate new information and converge slowly.

Similarly, if the discount rate is too high, the agent may prioritise long-term rewards too heavily, leading to slow convergence as it takes more time to explore the environment and learn. On the other hand, if the discount rate is too low, the agent may prioritise short-term rewards too heavily, leading to suboptimal behaviour as it may not consider the long-term consequences of its actions.

## Assumptions

- One assumption in this analysis is that the environment and the Q-learning algorithm are fixed, and only the values of alpha and gamma are changed.
- Assume that the sum of each episode shows the amount of new knowledge the agent learns in the environment.
- Another assumption is that the agent can learn and improve its performance over time, which may not always be the case if the learning process is heavily hindered by other factors such as lack of exploration or insufficient information.

## References

[1] S. Bhatt, "Reinforcement Learning 101," *Medium*, Apr. 19, 2019. https://towardsdatascience.com/reinforcement-learning-101-e24b50e1d292 (accessed Dec. 26, 2022).

[2] "Tutorial: An Introduction to Reinforcement Learning Using...," *Tutorial: An Introduction to Reinforcement Learning Using OpenAI Gym | Coder One*. https://www.gocoder.one/blog/rl-tutorial-with-openai-gym (accessed Dec. 26, 2022).