Sri Lanka Institute of Information Technology

**Deep Learning – SE4050**

Lab 08 – 2024, Year 4 Semester 1

A picture containing text, clipart, vector graphics

Description automatically generated

<<IT21308598 – Mahawaththa N.T.M.A.S.M>>

**Question 01**

1. Completed parts of the Markov\_Decision\_Process (PolicyIteration) notebook

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer program

Description automatically generated

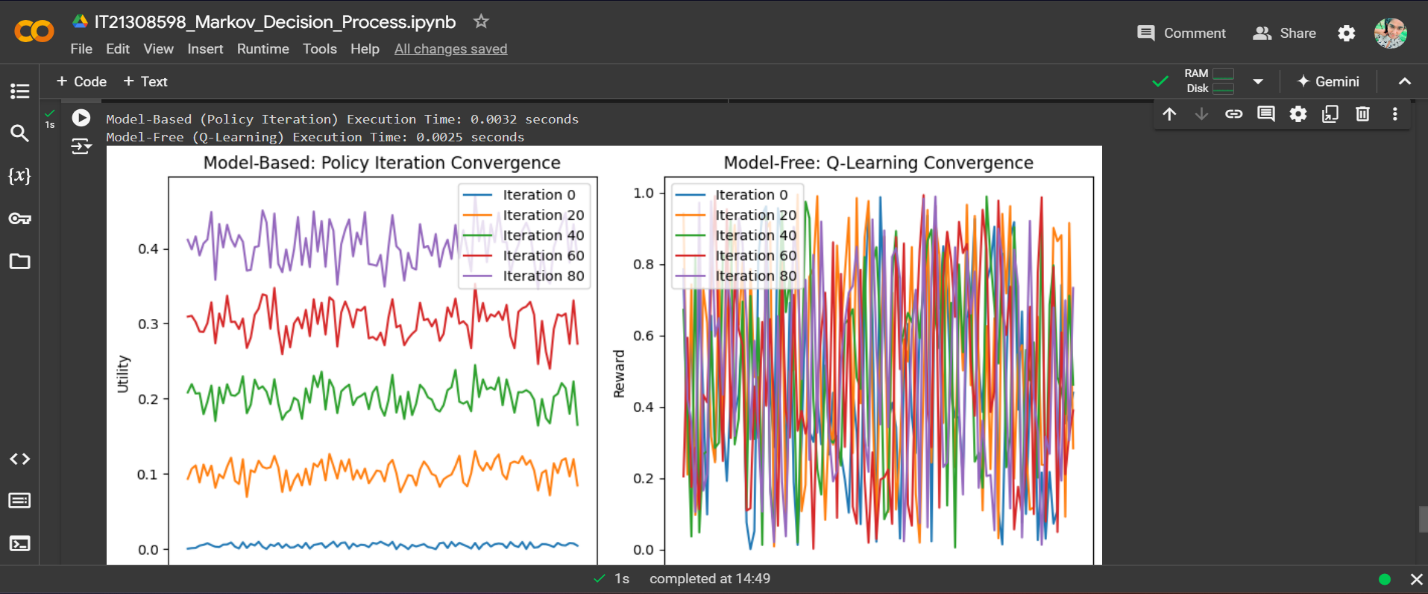
1. A screenshot of a computer program

   Description automatically generatedCompleted parts of the GridWorld (QLearning) notebook

**Question 02**

The difference between Model-Based and Model-Free algorithms briefly

* **Model-Based Algorithms (Policy Iteration, Value Iteration)**:
  + These algorithms require a model of the environment (i.e., transition probabilities and reward function).
  + They explicitly calculate the optimal policy by iterating over state transitions.
  + They tend to converge faster since they use the complete model of the environment.
* **Model-Free Algorithms (Q-Learning)**:
  + These algorithms do not require a model of the environment.
  + They learn the optimal policy through exploration of the environment by interacting with it.
  + They generally take longer to converge, as they must explore the state-action space to estimate the optimal policy.



A screenshot of a computer

Description automatically generated

**Question 03**

* **Observations**
* In this task, we extended the Markov Decision Process (MDP) with Q-Learning to develop a Deep Q-Learning (DQN) model.
* The Q-values are approximated using a neural network instead of a traditional lookup table.
* We implemented an epsilon-greedy strategy to balance exploration and exploitation, and tested different epsilon values (0.1, 0.5, 0.9) to analyze their impact on the performance of the agent.
* The epsilon-greedy strategy plays a critical role in ensuring that the agent explores the environment sufficiently while still exploiting the knowledge it has gained so far.
* As we decrease the epsilon value, the agent becomes more exploitative, while larger epsilon values promote more exploration.
* The results show that moderate epsilon values (e.g., 0.5) tend to strike a balance between exploration and exploitation, leading to more consistent learning and faster convergence.

**The following plot illustrates the performance of DQN with different epsilon values over 50 episodes:**

A screenshot of a computer

Description automatically generated