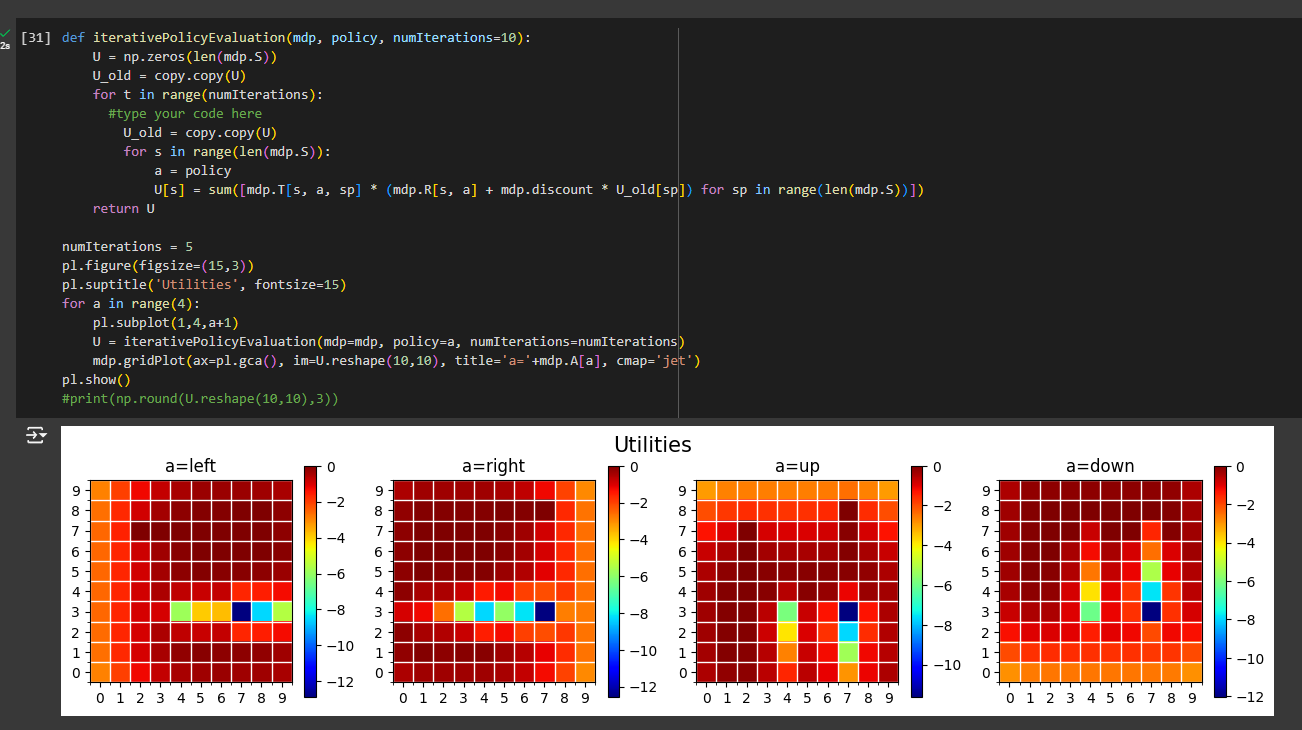
LAB 08

Question 1: Markov Decision Process and Q-Learning



A screenshot of a computer program

Description automatically generated

A computer screen with many colorful text

Description automatically generated

Question 2: Model-Based vs Model-Free Reinforcement Learning

A screen shot of a computer program

Description automatically generated

A computer screen shot of a program

Description automatically generated

**Model-Based Reinforcement Learning:**

* In **Model-Based RL**, the agent has access to a model of the environment, meaning it knows the transition probabilities (P(s'|s, a)) and the reward function (R(s, a)).
* The agent uses this model to plan its actions by calculating the best policy using algorithms like **Policy Iteration** or **Value Iteration**.
* **Advantages**:
  + Model-Based RL typically converges faster because it explicitly reasons about the environment.
  + It can be more sample-efficient as it plans over the known model.
* **Disadvantages**:
  + It requires knowledge of the environment’s model, which may not always be available.

**Model-Free Reinforcement Learning:**

* In **Model-Free RL**, the agent does not know the transition probabilities or reward function in advance.
* The agent learns by interacting with the environment, improving its estimates of the best actions directly through trial and error using algorithms like **Q-Learning**.
* **Advantages**:
  + Model-Free RL can be used when the environment’s model is unknown or too complex to model.
  + It is more flexible since it doesn’t need an explicit model of the environment.
* **Disadvantages**:
  + It usually requires more interactions with the environment (more data) and can take longer to converge.
  + It may be less sample-efficient as it learns from experience rather than planning over a model.

Question 3: Introduction to Deep Q-Learning (DQN)

1 A black screen with white lines

Description automatically generated

2

A screen shot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

 **Higher epsilon values** (e.g., 0.9) will lead to more exploration, which is useful in early stages of learning but may delay convergence.

 **Lower epsilon values** (e.g., 0.1) promote exploitation, which can lead to faster convergence if the model has already learned a good policy.

 A balance is necessary for optimal learning, and using an **epsilon decay** strategy helps shift from exploration to exploitation as the model improves.