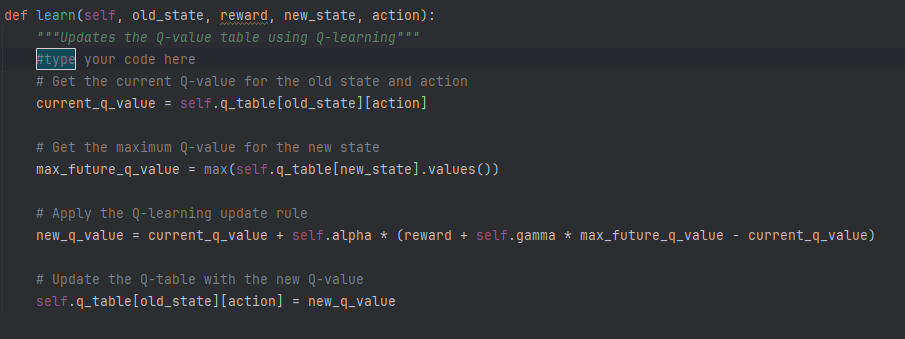
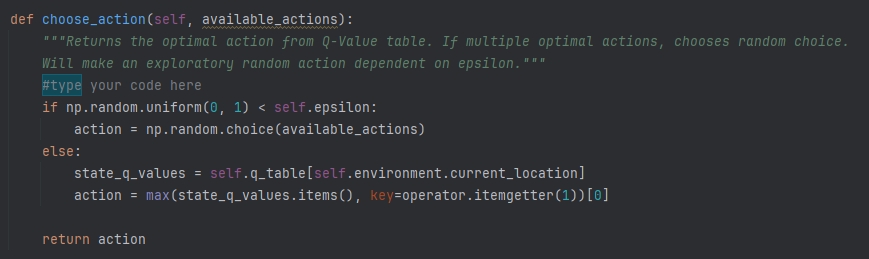
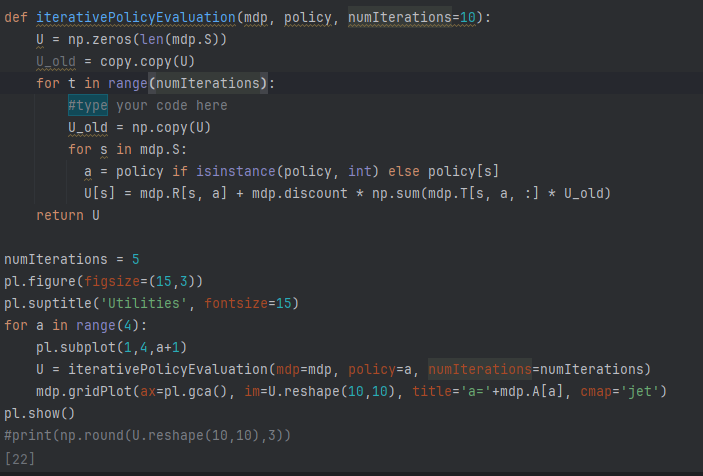
**Deep Learning - Lab Sheet 8**

1. **Markov Decision Process and Q-Learning**

A screenshot of a computer program

Description automatically generated

A screenshot of a computer program

Description automatically generated

1. **Model-Based vs Model-Free Reinforcement Learning**

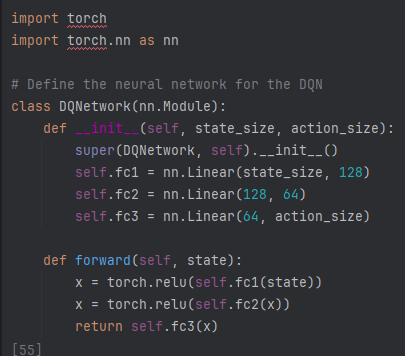
Model-Based algorithms build or use an explicit model of the environment, which includes the transition dynamics (how one state leads to another) and the reward structure. These algorithms use this model to plan their actions by predicting future outcomes based on potential decisions. Planning methods, such as dynamic programming or search algorithms, are typically used. Since they have access to a model, Model-Based methods can often make more informed decisions faster. However, they require more computational resources and might not be feasible when the environment is complex or hard to model. Examples include Value Iteration and Dyna-Q.

A screenshot of a computer screen

Description automatically generatedOn the other hand, Model-Free algorithms learn directly from experience, without constructing an explicit model of the environment. They rely on trial and error to understand which actions lead to higher rewards, either by learning the value of states or state-action pairs. These methods do not predict future states; instead, they focus on optimizing the policy based on the observed rewards and actions taken. While they are more straightforward and generally less computationally expensive, Model-Free algorithms may take longer to converge as they explore the environment. Examples include Q-Learning and Policy Gradient methods.

1. **Introduction to Deep Q-Learning (DQN)**

A screen shot of a computer program

Description automatically generated

A screen shot of a computer program

Description automatically generated

A graph of different colors and numbers

Description automatically generated with medium confidenceA computer screen shot of a program

Description automatically generated

**Observations:**

1. **Epsilon 0.9**

With a high epsilon (0.9), the DQN model is heavily biased towards exploration. The plot shows a lot of fluctuations in performance throughout the episodes, indicating that the model is trying out different actions frequently. However, around 70-100 episodes, there is a noticeable improvement in performance, as the model begins to balance between exploring new actions and exploiting learned strategies.

1. **Epsilon 0.5**

At epsilon 0.5, which represents a more balanced approach between exploration and exploitation, the model initially fluctuates but starts to show signs of improvement earlier than the epsilon 0.9 model. Around the 50th episode, there are significant spikes in performance, and the trend remains more stable compared to epsilon 0.9.

1. **Epsilon 0.1**

With epsilon set to 0.1, the model is more focused on exploitation, meaning it relies more on the learned values rather than exploring. The performance remains quite flat for most of the episodes, indicating that the model may not be exploring enough to discover better strategies or may have converged to suboptimal policies early on.