**Theory Question 2**

In your own words, compare and contrast policy iteration and value iteration in a few sentences. I.e. what are the advantages and disadvantages
of either? What are the types of problem to which they can be applied? Which is faster under which circumstances?

The aim of both these iterative processes is to get to an optimal policy for a given Markov Decision Process (MDP). The policy iteration uses the Bellman expectation equation and a greedy policy improvement to iteratively improve its policy. It reaches the optimal policy only once the estimates of state-value function converge to $\lim_{k\to\infty} v_k$ or at least until the difference between the old and new state-value function estimates becomes very small ($\Delta < \Theta$). This can mean several iterations where the optimal policy has already been achieved, but the estimates of the state-value function (output of the Bellman expectation equation) are still converging.

The value iteration on the other hand does not further iterate once the optimal policy has been achieved. It works by turning the Bellman optimality equation into an update rule. The iterations here are truncated until the Bellman optimality equation is reached for all state action pairs and we have an optimal policy.

The main advantage of the value iteration over the policy iteration is that it can converge in fewer steps. It stops when it has reached the optimal policy, while the policy iteration can continue iteration, although the optimal policy has been reached. Another advantage of the value iteration is that it only relies on the Bellman optimality equation rather than requiring both the Bellman expectation equation and a greedy policy improvement.

Both these algorithms can be applied to finite environments, where actions, states and rewards are finite too. Common applications are certain boardgames, video games, as well as control and navigation of robots that can be modelled as Markov Decision Processes.