**Experiment :- 10**

|  |
| --- |
| **Title:**  **Write a program to implement approximate value iteration (AVI) algorithm and API (Approximate value iteration)** |

# Objective:

# Books/ Journals/ Websites referred:

* Markov Decision Processes in Artificial Intelligence MDPs, Beyond MDPs and Applications, Edited by Olivier Sigaud, Olivier Buffet, Wiley Publications, 2010
* https://gym.openai.com/docs/
* https://www.analyticsvidhya.com/blog/2018/09/reinforcement-learning-model-based-planning-dynamic-programming/

# Resources used:

# Python

# Gym

# Theory:

# Approximate Policy Iteration (API) is a reinforcement learning algorithm that combines the concepts of policy evaluation and policy improvement in an iterative process to find an optimal policy for a given problem.

# The algorithm consists of the following steps:

# 1) Initialization: Initialize the policy and the value function approximator.

# 2) Policy Evaluation: Evaluate the value of the current policy by computing the expected return for each state under the policy. This is done by iterating over the state space and updating the value function approximator using the Bellman equation.

# 3) Policy Improvement: Improve the current policy by selecting the action that maximizes the expected return for each state. This is done by iterating over the state space and selecting the action that gives the highest Q-value.

# 4) Convergence: Repeat steps 2 and 3 until the policy and the value function approximator converge.

# The value function approximator can be any function approximator such as a neural network, linear function approximator, or any other function approximator.

# In each iteration, the policy evaluation step updates the value function approximator by minimizing the TD error between the current value estimate and the expected return using the current policy. The policy improvement step updates the policy by selecting the action that gives the highest Q-value.

# API has been shown to be effective in a wide range of applications, including robotics, game playing, and autonomous driving.

# Implementation (Code):

**Kindly find code and output in E10\_API.ipynb**

# Output Screenshots:

# Conclusion (Students should write in their own words):

# The Approximate Policy Iteration algorithm is a powerful technique for solving reinforcement learning problems with large state spaces. It combines the advantages of policy iteration and value iteration, making it a computationally efficient and effective algorithm. By alternating between policy evaluation and policy improvement, the algorithm converges to an optimal policy while updating the value function approximation in each iteration.

# One key advantage of the approximate policy iteration algorithm is that it can handle continuous state and action spaces, as well as high-dimensional state representations, making it applicable to a wide range of real-world problems. It also allows for the use of function approximation techniques, such as neural networks or linear models, to represent the value function and policy.

# Overall, the approximate policy iteration algorithm provides a powerful and flexible framework for solving reinforcement learning problems, making it an important tool for researchers and practitioners in the field.

# Applications:

# The Approximate Policy Iteration algorithm has several applications in the field of reinforcement learning and control. Some of its applications include:

# 1) Robotics: The Approximate Policy Iteration algorithm can be used to develop control policies for robots to perform various tasks. The algorithm can be used to learn the optimal control policy for a robot to perform a specific task, such as grasping objects or navigating in a complex environment.

# 2) Gaming: The Approximate Policy Iteration algorithm can be used to develop intelligent agents for playing games. The algorithm can be used to learn the optimal policy for an agent to take actions in a game to maximize its reward, such as winning the game or achieving a high score.

# 3) Finance: The Approximate Policy Iteration algorithm can be used in finance to develop trading strategies. The algorithm can be used to learn the optimal policy for a trader to take actions, such as buying or selling stocks, to maximize their profit.

# 4) Healthcare: The Approximate Policy Iteration algorithm can be used in healthcare to develop personalized treatment plans for patients. The algorithm can be used to learn the optimal treatment policy for a patient based on their medical history and current condition.

# Overall, the Approximate Policy Iteration algorithm is a powerful tool for developing optimal control policies in various applications, where traditional optimization techniques may not be feasible or efficient.