F29AI

CourseWork 2: MDPs & Reinforcement Learning; Tic Tac Toe

Policy Iteration

Question 4 (1 point): As in Question 2, this time test your Policy Iteration Agent against each of the provided agents 50 times and report on the results – how many games they won, lost & drew. The other agents are: *random*, *aggressive*, *defensive*.

This should take the form of a very short .pdf report named: `pi-agent-report.pdf`. Commit this together with your code, and push to your fork.

Against Defensive Agent:

Wins: 45 Losses: 0 Draws: 5

Against Aggressive Agent:

Wins: 49 Losses: 0 Draws: 1

Against Random Agent:

Wins: 49 Losses: 0 Draws: 1

initRandomPolicy(): This method initializes a random policy by assigning a random valid move to each non-terminal game state. It ensures every state has an initial action, even if arbitrary, to begin the policy iteration process.

evaluatePolicy(double delta): This method iteratively updates the policy values for each state by calculating the expected value of following the current policy until the value changes are smaller than the delta threshold, ensuring convergence.

improvePolicy(): This method refines the current policy by evaluating all possible moves for each state, selecting the move that maximizes the expected value, and updates the policy if a better move is found.

train(): This method alternates between evaluating the current policy and improving it until the policy stabilizes, then converts the final refined policy into a usable Policy object for decision-making.