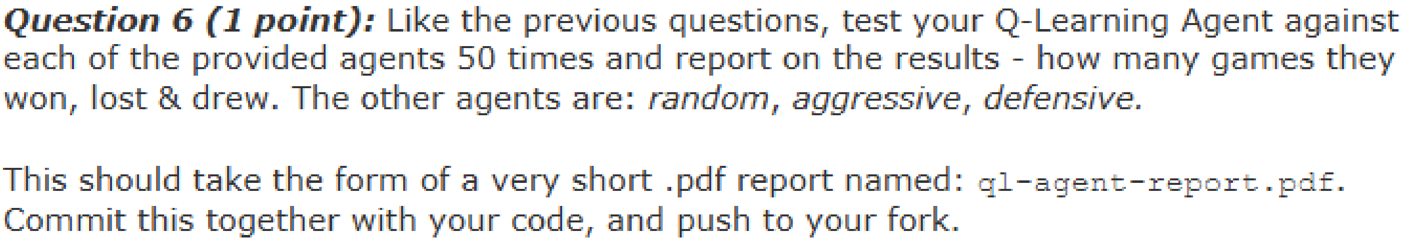
F29AI

CourseWork 2: MDPs & Reinforcement Learning; Tic Tac Toe

Q-Learning



Question 6 (1 point): Like the previous questions, test your Q-Learning 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: `ql-agent-report.pdf`. Commit this together with your code, and push to your fork.

Against Defensive Agent:

Wins: 42 Losses: 0 Draws: 8

Against Aggressive Agent:

Wins: 49 Losses: 0 Draws: 1

Against Random Agent:

Wins: 49 Losses: 0 Draws: 1

train():

This method runs multiple episodes where the agent interacts with the environment. For each state, the agent selects a move based on an epsilon-greedy policy (explore or exploit), executes the move, and updates the Q-value using the Q-learning formula based on the reward and the maximum Q-value of the next state. This process is repeated until the environment reaches a terminal state for each episode.

extractPolicy():

This method generates the agent’s policy by iterating over all non-terminal states and selecting the move with the highest Q-value for each state, effectively mapping each state to its optimal action according to the learned Q-values. This policy guides the agent's behavior once training is complete.