**Pathfinding with Reinforcement Learning**

**Path Attained**

|  |  |
| --- | --- |
|  | Paths Obtained |
| Q Learning | 10 |
| SARSA | 10 |
| Actor-Critic (β=0.3) | 7 |

For the Actor-Critic method, modification in parameters was needed to consistently create a policy that produces an optimal path to goal. A sweeping function was made to sweep β from 0 to 1 to find the β value that produces the most paths out of 10 runs.

The final β value was set to 0.3, as this performed the best.

**Optimal Policy + Optimal Path**

|  |  |  |
| --- | --- | --- |
|  | Optimal Policy | Optimal Path |
| Q Learning | A game with black squares and red and blue squares  Description automatically generated with medium confidence | A game of crossword  Description automatically generated |
| SARSA | A game with a maze  Description automatically generated with medium confidence | A crossword puzzle with a maze  Description automatically generated |
| Actor Critic (β=0.3) | A game with black and red squares  Description automatically generated with medium confidence | A crossword puzzle with a maze  Description automatically generated |

**Average Accumulated Reward**

|  |  |
| --- | --- |
|  | Average Accumulated Reward |
| Q Learning | A graph of blue lines  Description automatically generated |
| SARSA | A graph of blue lines  Description automatically generated |
| Actor Critic  (β=0.3) | A blue lines with white text  Description automatically generated |

**All Accumulated Reward Comparison**

A graph of a number of green and blue lines

Description automatically generated

**Discussion:**

The results for Q-Learning and SARSA are very close, with the SARSA with a bit more variation. This is due to the SARSA target being less greedy than Q-Learning, giving it a bit more exploration.

The results for Actor-Critic varies greatly as it does not always find a path to goal and is much more inconsistent.

All algorithms converged quite quickly, it is difficult to see which one converged faster.