Evaluation of the Agent

To evaluate the agent, you typically consider metrics such as cumulative rewards, convergence rate, and policy quality. Here is how we can discuss the results in relation to the FrozenLake-v1 environment and earlier questions:

1. Cumulative Rewards

Observation: Assess how much reward the agent accumulates over a number of episodes. A higher cumulative reward indicates that the agent is effectively learning to navigate the frozen lake and avoid holes.

Discussion: Compare this to baseline performance (random actions or a simpler policy). If the agent outperforms random behavior, it confirms that your learning algorithm is successfully capturing the environment’s dynamics.

2. Convergence Rate

Observation: Analyze how quickly the agent's learning stabilizes. Faster convergence suggests an efficient learning process.

Discussion: Relate this to the exploration-exploitation strategy, if for example an epsilon-greedy policy was used, or the hyperparameters such as learning rate and discount factor. Where these optimized to facilitate convergence?

3. Policy Quality

Observation: Assess the optimality of the policy by visualizing the learned policy or simulate it in the environment.

Discussion: Check whether the agent consistently finds the shortest path to the goal or avoids falling into the holes. If not, analyze where the policy struggles, for example, in stochastic transitions or slippery tiles.

4. Comparison with Previous Questions

Relate the results back to 1.1 to 1.3, where the environment dynamics were discussed. For instance:

Does the agent perform well under stochastic conditions, or does it struggle with variability in outcomes?

Which states are harder, and how does that correspond to your theoretical expectations based on the environment of?

Highlight if the choice of algorithm (Q-learning/ SARSA) corresponds to the ability of the agent to adapt to the dynamics of FrozenLake-v1.

5. Performance under Deterministic vs. Stochastic Settings

If the agent was tested in both settings:

Deterministic: Will most probably yield better performance since it can depend on predictable outcomes.

Stochastic: The performance will deteriorate; it reflects added complexity due to uncertainty.

Conclusion

Summarize whether the agent’s performance aligns with expectations.

Identify areas for improvement, such as:

Adjusting hyperparameters.

Enhancing exploration strategies.

Using advanced methods like Deep Q-Networks (DQN) for better generalization.