Q-Learning on FrozenLake-v1: Effects of α, γ, and ε on Learning

1. Environment and Setup
Environment: FrozenLake-v1 (4x4, is_slippery=True).
States: 16 tiles (S, F, H, G). Actions: {Left, Down, Right, Up}.
Reward: 0 per step, 1 for reaching the goal, episode ends on hole or goal.
Episodes: 5000, averaged across 5 seeds. ε decays from 1.0 → 0.05 with rate 0.9995.
Exploration–exploitation controlled via ε-greedy policy.

2. Algorithm: Q-Learning
Update rule:
$Q(s,a) \leftarrow Q(s,a) + \alpha [r + \gamma \max_{a'} Q(s',a') - Q(s,a)]$
Exploration: ε-greedy, where random actions occur with probability ε.
Learning parameters: $\alpha \in \{0.1, 0.3, 0.5\}$, $\gamma \in \{0.9, 0.99\}$.

3. Results and Observations
Higher α led to faster learning but less stable convergence.
γ = 0.99 outperformed γ = 0.9, showing that valuing future rewards helped in sparse reward settings.
Slower ε decay improved long-term performance by maintaining exploration longer.
Below is a typical learning curve showing moving-average reward over episodes.

(Insert learning curve figure here: qlearning_frozenlake_curves.png)

4. Discussion
• α (learning rate): Too low → slow progress; too high → oscillations.
• γ (discount factor): Higher γ improved stability and long-term goal pursuit.
• ε (exploration): Gradual decay ensured sufficient exploration before exploitation.
Overall, balanced parameter choices (α=0.3, γ=0.99, ε_decay=0.9995) yielded the best results.

5. Reflection
The experiment demonstrated the exploration–exploitation dilemma clearly. Excessive exploration delayed convergence, while premature exploitation trapped the agent in suboptimal policies. Future work could include comparing SARSA with Q-learning and testing on larger maps (e.g., 8x8 FrozenLake). **References:**
- OpenAI Gym Documentation: https://www.gymlibrary.dev/environments/toy_text/frozen_lake/
- Towards Data Science Q-Learning Tutorial: https://towardsdatascience.com/simple-reinforcement-learning-q-learning-fcddc4b6fe56