Department of Computer Science
University of Cape Town, South Africa

CSC3022F Machine Learning

Assignment 2 - Reinforcement Learning

Prosper Arineitwe Asiimwe - ARNARI002

May, 2025

1 Scenario 1: Exploration strategy comparison report

We compared ϵ-greedy and Softmax exploration strategies across different hyperparameters. The best performance was achieved by the softmax, alpha: 0.05, gamma: 0.99, epsilon decay: 0.995 configuration. Softmax exploration demonstrated faster convergence and more stable rewards, with smoother learning curves across episodes. In contrast, ϵ-greedy often required more episodes to stabilize and showed higher variance. The flexibility of Softmax to adjust action probabilities enabled more effective exploration, resulting in a more consistent final policy. This illustrates that in sparse reward environments like the Four-Rooms domain, softmax can outperform ϵ-greedy by maintaining a better balance between exploration and exploitation.

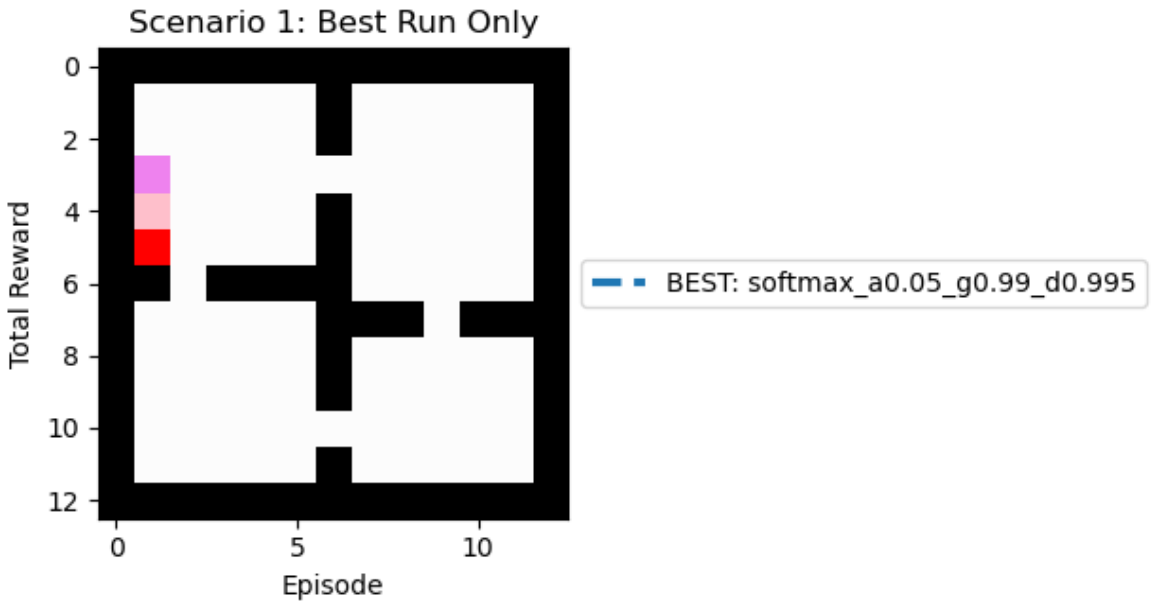

Figure 1: The best performance achieved by the softmax, alpha: 0.05, gamma: 0.99, epsilon decay: 0.995 configuration.