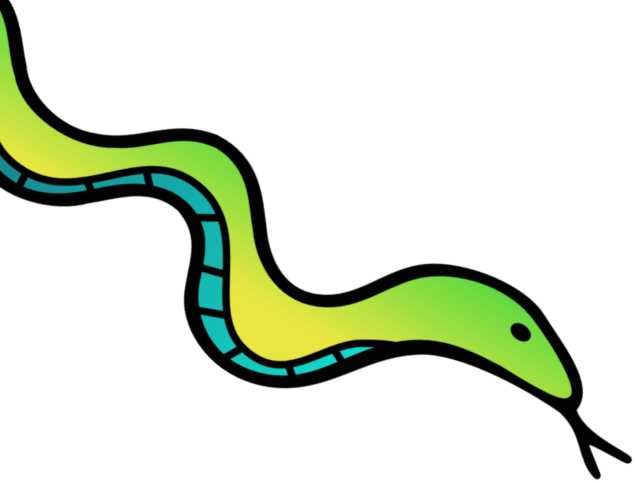


Naggin' Naagin

Self-Learning Snake

CS551: Introduction to Deep Learning
Course Project

Don Kurian Dennis - 1301CS17

Ritobroto Maitra - 1301CS50

Outline

- Introduction
- Recap
- Exact Q-Learning
- Approximate Q-Learning
- Results
- Observations
- Further Work

Introduction

- Modelling the Snake game has two aspects:
  - Game play with a world model (perfect or partial information)
  - Learning with the world model through training (no domain knowledge)

- Stage 1: Explore Different Search and MDPs
- Stage 2: Q-Learning with Compressed States
- Stage 3: Approximate Q-Learning with Features
- Stage 4: Higher Level Features from Deep Learning

Recap

- So far, we have explored:
  - Search-based approaches
  - Modelling using MDP
  - Reinforcement Learning using MDP
  - Value and Policy Iteration on known MDP

Video Placeholder

Exact Q-Learning

- Due to the large state space, we compressed it into 4 quadrants relative to the snake's head.

- Very simple model; converges to the policy of always moving diagonally towards the food.

Video Placeholder

Approximate Q-Learning

- Instead of using states; we use a linear combination of features using weights in Approximate Q-Learning
- During training, it is similar to a function approximation except that we use a control-feedback loop.
- After training, the agent reduces to a reflex agent with a fixed policy.

Approximate Q-Learning

- Convergence of values (weights) guaranteed on infinite exploration
- We can't do infinite iterations so we use an epsilon-greedy strategy for an exploration vs exploitation strategy.
- Policy converges much before values do.

Approximate Q-Learning

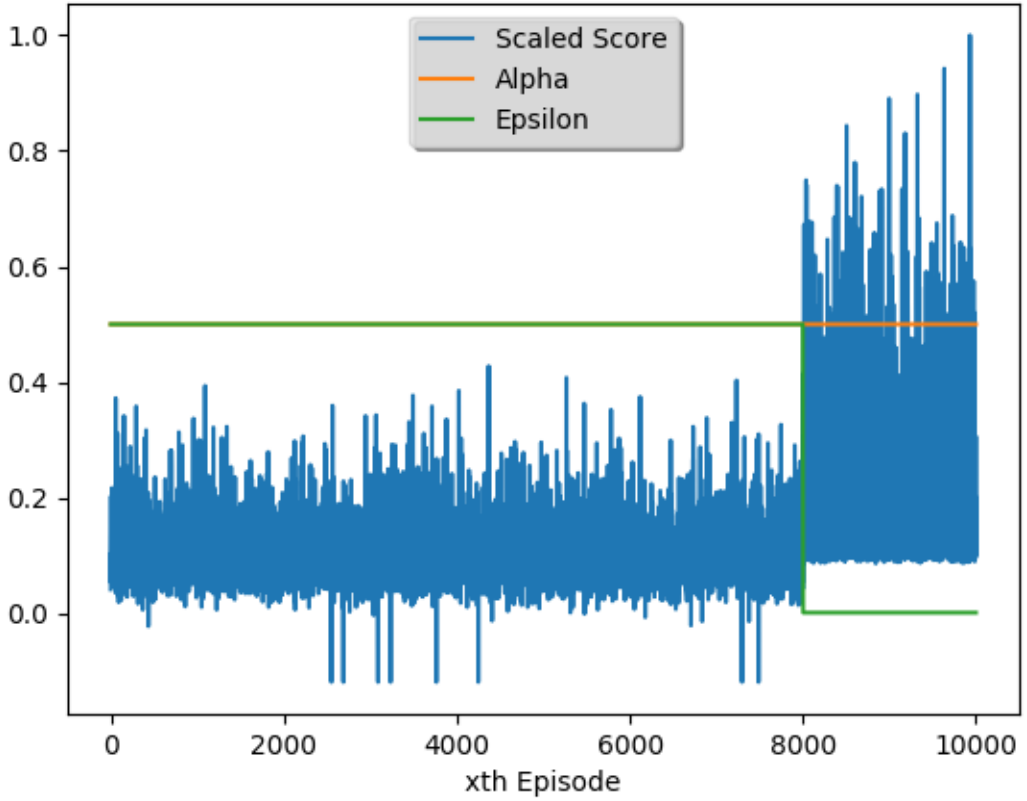
- Features used:
  - Circular Food Vicinity: Measures the circular manhattan distance of the food from the snake's head
  - Gradient Factor: Measures the likelihood of the snake of running into itself circularly in the direction of movement
  - Collision Factor: Measures the likelihood of the snake of colliding in general

Results

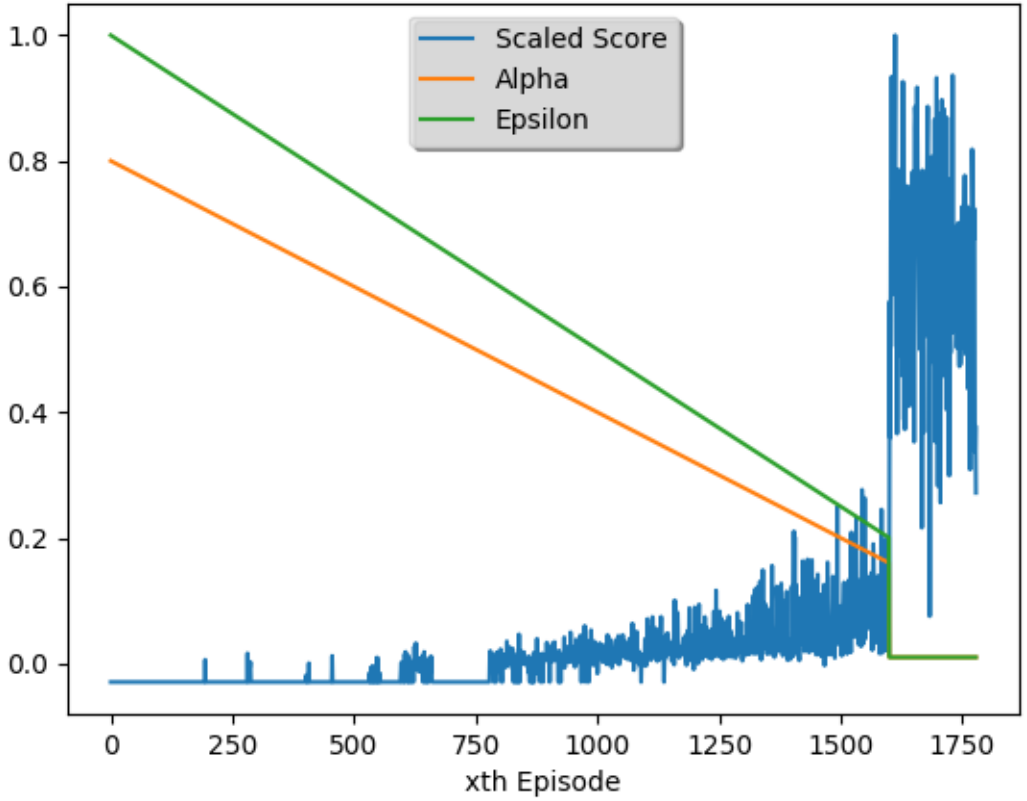
- Snake reaches a mean score of 2177 with a mean snake length of 65 (vs 1645 with a reflex agent and 1857 with a MinMax) in 20x20 grids
- Often reaches lengths > 100
- Covers nearly the entire grid for smaller grids (80% for 5x5)

Video Placeholder

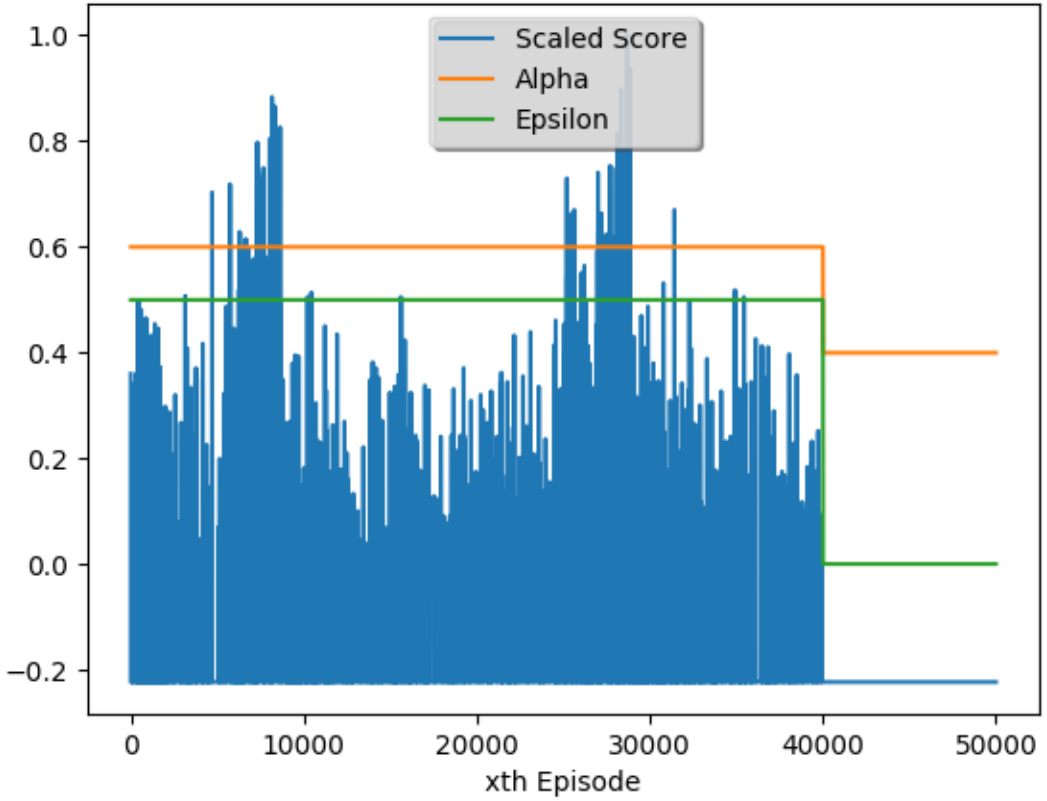
Observations



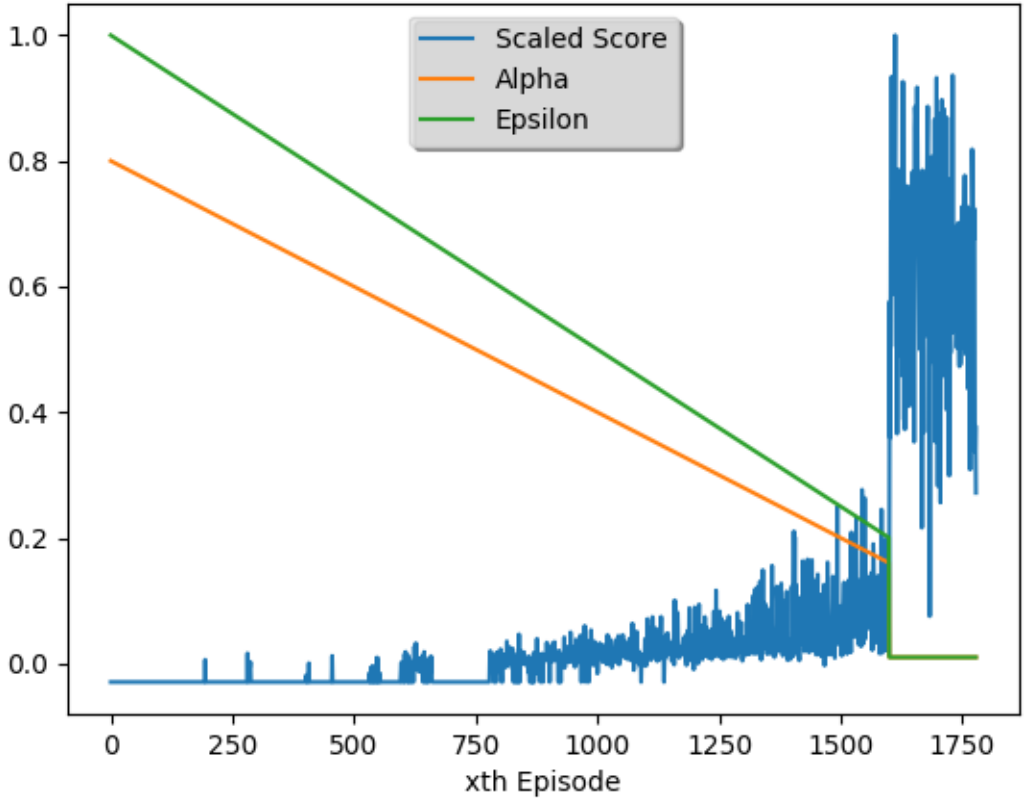
Observations



Observations



Observations



Observations

- Unbounded growth in values under certain conditions
- Unexplained thrashing
- Extremely correlated weights with a difference of exactly 1.000000
- Proof of convergence not formalized
- Scaling issues

Further Work

- Explain and rationalize observations
- Provide theoretical framework for explanations
- Explore Deep Q-Learning approaches for extracting higher-level features from sensory inputs

Thanks!