# Project 2: Reinforcement Learning - Q-Learning and Deep Q-Learning

Alexander Svarfdal Gudmundsson Jan Babin

November 1, 2024

# Contents

# 1 Introduction

The main objective of this project is to apply Q-learning and Deep Q-learning to train an agent in a Flappy Bird game. The reason why reinforcement learning is applicable is because we can make an agent run in an simulated environment, even making the agent play at a much faster pace than in real time. In Flappy Bird you as the player play as a bird in a 2d environment where there are two pipes that appear to the right of the screen one coming from above and one coming from the top, with each consecutive frame the pipes move closer to the bird, and your task is to locate the bird by flapping its wings to make it fit between these two pipes. Flappy Bird is an interesting game to take on because it has a good balance between simplicity and challenge. Q-learning algorithm uses a tabular structure to learn every single possible state of the game to then take the best possible action that it knows. The deep-Q learning uses function approximation to estimate whats the best action depending on what state the game is in.

# 2 Problem Description

Describe the problem, including the specific objectives: achieving the best policy and learning a policy as quickly as possible. Discuss any trade-offs involved in the process.

# 3 Characterization of the Environment

Discuss the environment characteristics (discrete/continuous, finite/infinite, episodic/continuous), what constitutes a state, and whether it is a Markov process. Mention the feasibility of using Q-Learning or Deep Q-Learning and the challenges expected.

# 4 Implementation of Q-Learning

## 4.1 State Representation and Action Space

Explain the state representation used (e.g., player_y, next_pipe_top_y, next_pipe_dist_to_player, and player_vel) and the action space (flap or do nothing).

- **player_y**: The current vertical position of the bird.
- **next_pipe_top_y**: The position of the next top pipe.
- **next_pipe_dist_to_player**:
- **player_vel**:

## 4.2 Q-Learning Algorithm

Provide details on the discretization of the state space, $\epsilon$-greedy policy, learning rate, and update rule. Mention how terminal states are handled.

# 5 Learning Curve Analysis

Include a plot of the learning curve and interpret the results. Discuss what the curve reveals about the performance and stability of the Q-Learning algorithm.

# 6 Implementation of Deep Q-Learning

## 6.1 Model Setup

Explain the structure of the neural network, including the input normalization, hidden layers, output layer, and activation functions.

## 6.2 Algorithm Details

Discuss the modifications for Deep Q-Learning, including the use of experience replay and the target network. Provide information about the update procedure and batch training.

## 6.3 Training and Evaluation

Explain how the training was conducted and include results with an analysis of the performance compared to the tabular Q-Learning approach.

# 7 Experimental Results and Analysis

## 7.1 Parameter Tuning

Describe the experiments conducted with different parameter configurations (e.g., learning rate, $\epsilon$, batch size). Include a discussion on how these parameters influenced learning speed and policy quality.

## 7.2 Best Performing Agent

Report on the best results obtained and analyze why this setup was effective.

# 8 Conclusion

Summarize the findings of the project, the parameters that influenced learning the most, and potential future improvements.

# 9 Discussion and Future Work

# Appendix

| **Feature** | **Description** |
|---|---|
| *Age* | Coded in 13 age groups (e.g., 1: 18-24, 2: 25-29, etc.) |
| *Sex* | Sex of the individual (0: Male, 1: Female) |
| *HighChol* | High cholesterol (0: No, 1: Yes) |
| *CholCheck* | Checked cholesterol in the last 5 years (0: No, 1: Yes) |
| *BMI* | Body Mass Index (continuous variable) |
| *Smoker* | Smoked at least 100 cigarettes in their lifetime (0: No, 1: Yes) |
| *HeartDiseaseorAttack* | History of coronary heart disease or myocardial infarction (0: No, 1: Yes) |
| *PhysActivity* | Engaged in physical activity in the past 30 days, excluding work (0: No, 1: Yes) |
| *Fruits* | Consumes fruit 1 or more times per day (0: No, 1: Yes) |
| *Veggies* | Consumes vegetables 1 or more times per day (0: No, 1: Yes) |
| *HvyAlcoholConsump* | Heavy alcohol consumption (men: 14+ drinks/week, women: 7+ drinks/week) (0: No, 1: Yes) |
| *GenHlth* | Self-reported general health (1: Excellent, 2: Very good, 3: Good, 4: Fair, 5: Poor) |
| *MentHlth* | Days of poor mental health in the past 30 days (0 to 30) |
| *PhysHlth* | Days of poor physical health in the past 30 days (0 to 30) |
| *DiffWalk* | Difficulty walking or climbing stairs (0: No, 1: Yes) |
| *Stroke* | History of stroke (0: No, 1: Yes) |
| *HighBP* | High blood pressure (0: No, 1: Yes) |
| *Diabetes* | Presence of diabetes (0: No, 1: Yes) |

Table 1: Full list of dataset features used in the analysis.