Q-Learning

Peter (Petr) Ladur

56445760

Chapter 1

Introduction

The aim of this project is to explore Q-Learning in the context of tic-tac-toe and other applications using python. Theoretical Q-Table was generated using the “minimax” algorithm, and was compared to Q-Tables generated through training against both, random and perfect opponent. How hyperparameters (α, τ) affect training results was investigated and an optimal function for hyperparameters was proposed. Deep-Q-Learning was investigated in the context of different games in the “Open-AI gymnasium”.

1.1 Q-Table

Chapter 2

Background

2.1 Q-Table

Q-Table is a table which contains contains expected outcomes after every possible actions which can be made in a particular state. In the context of tic-tac-toe the state is the game position at that particular moment, and actions are all the empty squares. As seen in table 2.1, for that particualr state there are two actions that lead to a win and three actions which lead to a draw (assuming perfect play).

	(1, 1)	(1, 2)	(1, 3)	(2, 1)	(2, 2)	(2, 3)	(3, 1)	(3, 2)	(3, 3)
_ _ o _ x _ x o _		1	1			0	0	0	

Table 2.1: Two actions lead to a win, three actions lead to a draw assuming perfect play from both sides

Chapter 3

Q-Learning on tic-tac-toe

3.1 Generating perfect theoretical Q-Table

The “minimax” algorithm can be used in turn-based games designed to minimise the potential loss and maximise the potential gain. The algorithm works by first generating all the possible final states and then taking turns “undoing” the moves that could have lead to that state by removing **X’s**. If the result

3.2 Optimal and non optimal opponent

3.3 Variations in α and τ

In general the Q-Learning governing equation has 4 hyperparameters (α, τ, r and γ). The immidiate reward r is by default set at -1 for loss, 0 for draw, 1 for win and 0 for just making a move. γ is irrelevant as the future state matters as much as the current state (it doesn’t matter if naughts loose on move 5 or on move 7, they still lost).

However, τ and α matter significantly for the learning rate.

Chapter 4

Deep Q-Learning

Chapter 5

Conclusion

Chapter 6

Bibliography