MACHINE DATA AND LEARNING

REPORT ON ASSIGNMENT 2-PART 2

- Giri Prasath - 2018111006
- Shreevignesh - 2018111019

Introduction

Aim: Code the value iteration algorithm for given scenarios to obtain the optimal policy and the state reward values corresponding to it.

Task 1:

The code converges after 110 iterations of the algorithm choosing the best action for every iteration such that it has the maximum value each time.

In the first few iterations “SHOOT” was the most optimal in 6-8 states. This increases as the iterations increase and converges to 15 as the iteration goes to 110.

Similarly, “DODGE” was the most optimal for more than 24 states in the first few iterations and it decreases gradually and converges at 10 in the 110th iteration

Task 2:

Part 1:

The step cost reduces to -0.25. As the step cost has reduced the number of “SHOOT” action per iteration increases. This is because the penalty is associated with the “SHOOT” action and as it has reduced Lero would prefer “SHOOT” over “DODGE” in some states as the step-cost has reduced.

Part 2:

The gamma value has reduced from 0.99 to 0.1and the step cost is changed to -2.5. The behavior of Lero is weird in this case because the number of iterations has drastically dropped to 4 from more than a hundred. This is because in the previous case

the gamma was 0.99 thus we had several iterations but as gamma tends to 0 the algorithm becomes short-sighted and only wants the immediate reward as rewards in the future are worth much less than immediate rewards

Part 3:

The gamma and step cost remain 0.1 and -2.5 respectively but we change the delta value to 10^{-10} Delta being the Bellman-factor of error. As Delta reduces the number of iterations increases thus in this case too as the Bellman factor reduces from 10^{-3} to 10^{-10} as the number of iterations follows the formula

|Ui+1 - Ui | <= δ

It increases to reach convergence from 4 to 11.