

# CSC3022H – Q-Learning

## Philip Amwata

## Reward Function

### Movement Without Collision

Reward = -1 for movement without collision to discourage the agent from moving pointlessly around the environment.

### Collecting a Mine

Reward = 100 for collecting a mine to encourage the agent to move towards mines and collide with them.

### Collecting all Mines

Reward = 2000 for collecting all mines in the environment to encourage the agent to keep collecting mines.

### Collecting Super Mines

Reward = -200 to discourage the agent from moving towards super mines and collecting them.

## Decaying Learning Rate

Learning rate set to decay to decrease the weighting of present value.

## Discount Factor

A discount factor of 0.8 was selected to increase the weighting of future actions to encourage the bot to collect all mines in the environment.

## Results

Test Case	Number of Mines	Number of Super Mines	Average Number of Mines Collected	Maximum Number of Mines Collected
1	30	0	10.8	21
2	25	5	9.88	16
3	5	25	1.52	3

## Conclusion

In an environment where the agent can learn without dying too often (One with fewer super mines) the agent is able to learn to collect the mines. If the environment is filled with super mines, then the agent is unable to exploit

knowledge gained in the exploration phase as it is destroyed before completely exploring the environment.