Mandisa Baleni

BLNMAN002

11/Oct/2019

ML Assignment: Q-Learning

Results table over 50 runs

Test no.	Maximum no. of mines swept	Ave no. of mines swept
1	10	3.3
2	7	2.58
3	2	0.36

A learning rate of 0.6 was chosen. This is because a learning rate n needs to be greater or equal to 0 and less than or equal to 1 (0<=n<=1). Moreover, the closer a learning rate is to 1, the more the information acquired overrides old information. Setting a value to close to 1 would discount old useful information to 0 and setting a value to close to 0 would make the agent learn at a slower rate hence a moderate figure more in favor of new information it gathers. The chosen learning rate would influence the mine sweepers to determine how to collect the maximum number of mines by using old and new information proportionally but more in favor of new information. A bigger learning rate causes maximum and ave no of mines swept to increase but decreases number of sweeper deaths.

A discount factor of 0.6 was also chosen. The discount factor (gamda) also needs to be a factor that lies between 0 and 1 inclusive (0<=gamda<=1). A discount factor close to 1 will make the agent venture for a long-term high reward as opposed to short-term hence choosing 0.6 closer to 1 than 0. The choice to not choose this value too close to 1 is so that immediate rewards are not to undervalued. This will help using current information on rewards moderately whilst still considering future rewards to a larger extent. Increasing the discount factor makes the maximum and ave no of mines swept bellcurve with the best values lying in the centre.