Pratical session

Reinforcement Learning in a Discrete Domain

QUESTION 1

You get to visit a gold mine and you are allowed to spend as much time in it as you wish to get as much gold as possible. Working and spending time in the mine is complicated and rather than maximizing the quantity of gold mined, you want to optimize your satisfaction. You don't plan to stay in the mine for very long. Since you have followed the RL course last year, you have computed a discount factor of 95% which allows you to not spend too much time in the gold mine.

You start at the entry of the mine, at the ground level. You can decide to either go one level down, stay where you are or go one level up. To get to the gold, you have to dig trough several layers of soil. Staying at ground level, does not give you any satisfaction. The first two layers cost you one point of satisfaction each. The third layer contains a little gold and rewards you with one satisfaction point. If you decide to dig deeper, the next 4 layers are more complicated. Each layer further would respectively cost you 1,2,3 and 4 satisfaction points. However, the last layer is full of gold and digging there rewards you with 10 satisfaction points. If you decide to stay on a layer, you keep digging the same layer and it provides the same satisfaction over and over again. Provide a formalisation of this problem as seen in the theoretical course.

QUESTION 2

In the previous question, we add that at any point in time, the current mine layer can fall and force you to go back to the previous layer. The probability of such failure is 10%. Adapt the formalisation accordingly.

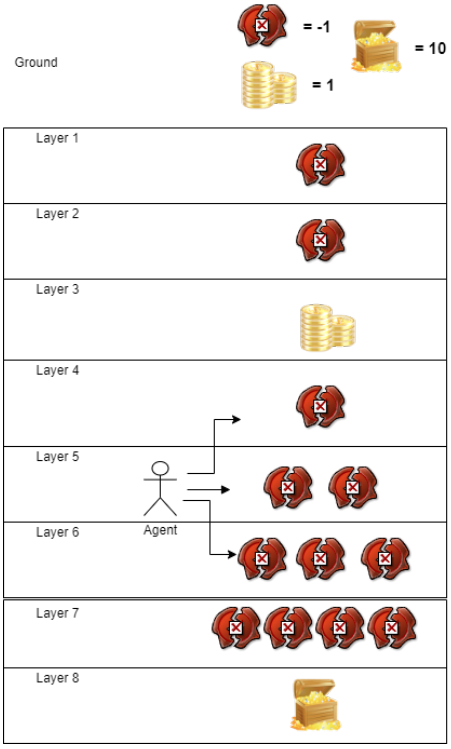

Figure 1: Illustration of the gold miner problem

QUESTION 3

Compute the expected cumulative reward over 8 first timesteps for an "always dig" policy in the deterministic case.

QUESTION 4

Compute the expected cumulative reward over 2 first timesteps for an "always dig" policy in the stochastic case.