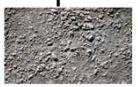
Exercise 5

Build the optimal policy for the Mineral explorer with Python

1. Follow the rule of taking a random first step and then following the optimal policy.
2. An episode is defined a – from start till the terminal state is reached.
3. Use the Bellman's equation
4. Implement the Q-table using the Bellman's equation and find the final optimal Q values.
5. Let gamma= 0.9 and alpha= 0.1.
6. Run for n episodes till the Q-values do not change anymore (optimal).
7. Print the final Q-values

How to verify your code

100  100	50 12.5 0	25 6.25 0	12.5 10 0	6.25 20 0	40  40

- We arrived at these Q-values by taking gamma = 0.5 and we did those computations manually.
- We did not use the updated Bellman equation, in other words, we assumed alpha = 1.
- So, if we coded the solution for Exercise 5 properly, and plugged in these value for alpha and gamma, then we should be able to verify this outcome and hence your code.