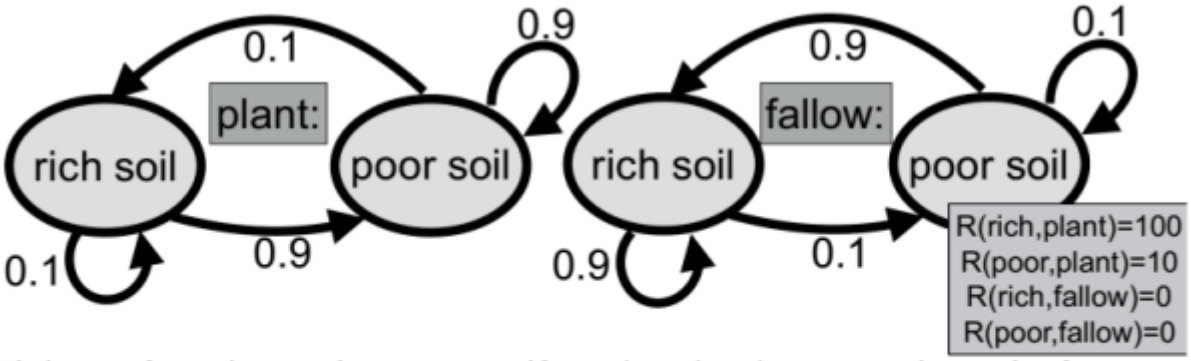
Practical - 7

Program Definition:

Write a python program to implement Q-value iteration algorithm for solving the Markov Decision Process (MDP) problem in the farmer example.



Value of action *a* in state *s* if make the best actions in future

$$Q^*(s,a) = R(s,a) + \gamma \sum_{s'} T(s,a,s') \max_{a'} Q^*(s',a')$$

What's the best policy?

$$\pi_{Q^*}(s) = \arg\max_a Q^*(s,a)$$

Output:

```
21012021001_ADESHARA BRIJESH

Q-values:
Q(rich, plant) = 529.02
Q(rich, fallow) = 470.89
Q(poor, plant) = 439.03
Q(poor, fallow) = 470.89

Optimal policy:
State: rich, Action: plant
State: poor, Action: fallow
```