Autonomous Orchard

An AI controlled orchard needs to decide when to harvest its trees. To do this it measures the concentration of three chemicals in the air. Each day the orchard can choose to wait or harvest. Waiting costs one credit in operating costs while a harvest ends the process. Once a crop is harvested, packaged and sold, the orchard is told the profit or loss of that harvest. Most experts agree that the function mapping the chemical concentrations to the profit is linear with some error.

1. The orchard has several samples of the profits from other harvests.

Concentration of A (ppm)	Concentration of B (ppm)	Concentration of C (ppm)	Profit/Reward (credits)
4	7	1	3
10	6	0	-15
20	1	15	5
4	19	3	21

Begin to approximate the function that maps the state feature vector to Q(state, harvest) using an MC goal. Do a gradient decent step on each sample. A sensible learning rate would be around 0.01, but feel free to try any value.

2. The orchard also has a record from its own last harvest.

Concentration of A (ppm)	Concentration of B (ppm)	Concentration of C (ppm)	Action	Profit/Reward (credits)
6	7	2	Wait	-1
3	8	4	Harvest	19

Continue to approximate Q(state, harvest) and start approximating Q(state, wait). Run through this episode, performing gradient decent and using the TD(0) goal. Instead of choosing your actions in an ε-greedy way, assume the ε-greedy algorithm chose the actions in the table above.

3. The orchard enters a new harvest season. Over the next three days the concentrations will be:

Day	Concentration of A (ppm)	Concentration of B (ppm)	Concentration of C (ppm)
1	20	6	1
2	10	7	2
3	5	8	4

Run through an episode using these values. Choose your actions greedily and don't update Q. If you wait on the third day then the episode is over. If your algorithm harvested, did it harvest at its maximum possible Q(s, harvest) or would it have had a higher value on one of the other two days? If your algorithm waited on the third day, which day was it closest to harvesting?

4. What is the overall effect of increasing the learning rate? What happens when you set it too high? What happens when you set it too low?
5. Write a short RL problem where it would be better to use feature vectors to represent a state. Write one where it would be better to use discrete states.