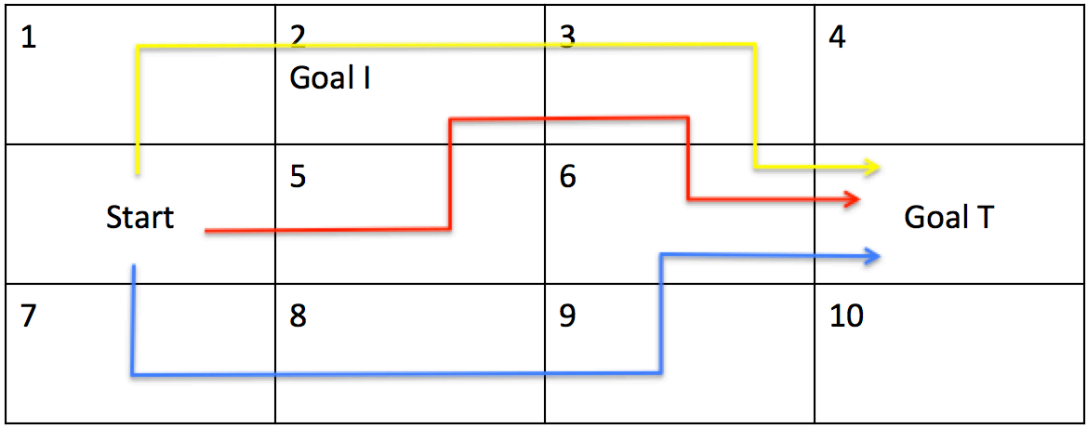
Reinforcement Learning

Suppose a reinforcement learning system is using Q-learning to learn how to navigate in an environment from a given start state. The environment has a *terminal* goal state (Goal T) that gives reward R(T)=10 and a *non-terminal* intermediate goal state (Goal I) that gives reward R(I) = 2.



Q Table:

State	Up	Down	Left	Right
1	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0
7	0.0	0.0	0.0	0.0
8	0.0	0.0	0.0	0.0
9	0.0	0.0	0.0	0.0
10	0.0	0.0	0.0	0.0

The discount factor is 0.5. The learning rate is 0.1. The agent does not have uncertain actions.

The first trial is marked blue.

1. During the first trial, compute the Q-table row for state 6.

The second trial is marked red.

2. During the second trial, compute the Q-table row for state 5.
3. During the second trial, compute the Q-table row for state 3.
4. During the second trial, compute the Q-table row for state 6.

The third trial is marked yellow.

5. During the third trial, compute the Q-table row for state 1.
6. During the third trial, compute the Q-table row for state 2.
7. During the third trial, compute the Q-table row for state 3.
8. During the third trial, compute the Q-table row for state 6.