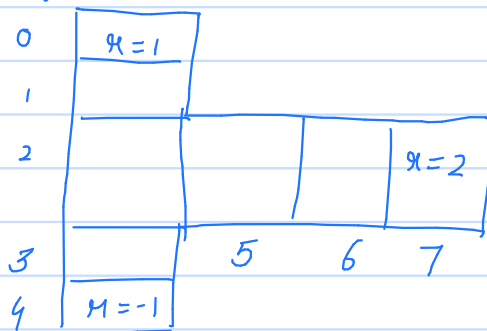8) Finding optimal Q values for T-maze.

States :- 0, 1, 2 ... 7.
action   Up (0) , down (1), Left (2), right (3)

R Matrix =		0	1	2	3
	0	—	0	—	—
	1	1	0	—	—
	2	0	0	—	0
	3	0	−1	—	—
	4	0	—	—	—
	5	—	—	0	0
	6	—	—	0	2
	7	—	—	0	—

Episode 1 :-
Consider the point at state $(s) = 2$

$Q(2, 0) = P(2, 0)$
$\qquad + \gamma(\max Q(1, 0), Q(1, 1))$
$\qquad = 0$

State 1 ,
$Q(1, 0) = P(1, 0) + \gamma \times (\max Q(0, 1))$
$\qquad = 1$

Episode - 2 :-

$Q(2, 3) = P(2, 3) + \gamma(\max(Q(5, 2),$
$\qquad\qquad\qquad\qquad Q(5, 3)))$
$\qquad = 0$

$Q(5, 3) = P(5, 3) + \gamma(\max(Q(6, 2), Q(6, 3)))$
$\qquad = 0$

$Q(6, 3) = P(6, 3) + \gamma(\max(Q(7, 2)))$
$\qquad = 2$

Episode - 3 :-

$Q(2, 1) = P(2, 1) + \gamma(\max(Q(3, 0), Q(3, 1))$
$\qquad = 0$
$Q(3, 1) = P(3, 1) + \gamma(\max(Q(4, 0)))$
$\qquad = -1$

Hence after traversing the environ-
ment until the Q values
are same
we finally get Q matrix as.

	0	1	2	3
0	—	0	—	—
1	1	0.25	—	—
2	0.5	0.125	—	0.5
3	0.25	-0.93	—	—
4	0.125	—	—	—
5	—	—	0.25	1
6	—	—	0.5	2
7	—	—	0	—