Reinforcement Learning - Homework 8
Vishal I B

The figure in 11.2 of the textbook, is an example of how the weights diverge when function approximation is applied to off-policy learning. Since the states are transitioned from lower to the higher values of weights, the TD error is calculated and the weights are increased. In this particular example, the value of W8 gets shot up as it is updated in every transition. Also, the target policy is always solid action, hence the update occurs over the solid action transition.

The feature matrix was defined as :

2	0	0	0	0	0	0	1
0	2	0	0	0	0	0	1
0	0	2	0	0	0	0	1
0	0	0	2	0	0	0	1
0	0	0	0	2	0	0	1
0	0	0	0	0	2	0	1
0	0	0	0	0	0	1	2

The weight vector was initialized to:

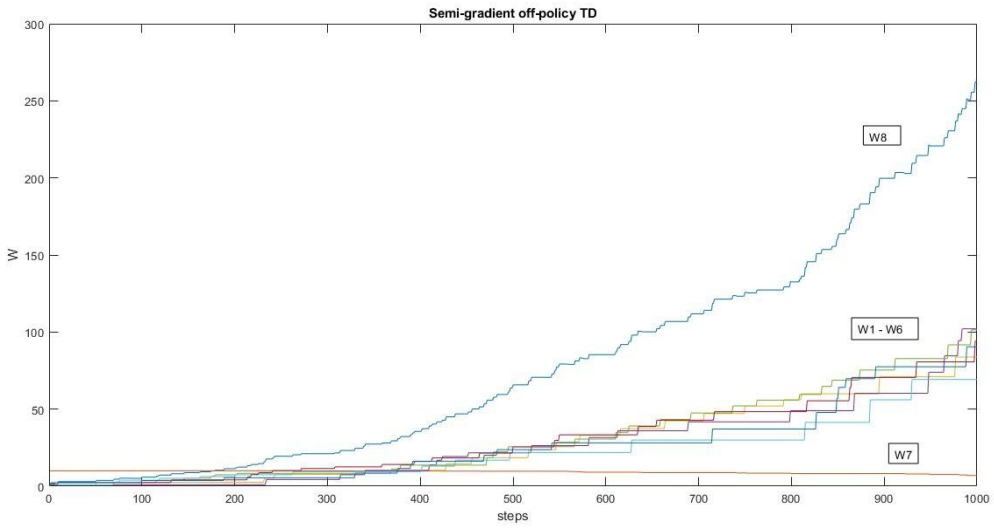
1	1	1	1	1	1	10	1

I ran the algorithm for 1000 steps and plotted the weights against the number of steps.

In the lecture, the professor said that the dashed action takes the next state to one of the 6 remaining states, but the book says that the dashed action takes the next state to one of the top 6 states (excludes the 7th state).

The figure for the second case seems closer to the one in the textbook. I have attached both the figures below.

The figure for the first case:



The figure for the second case: