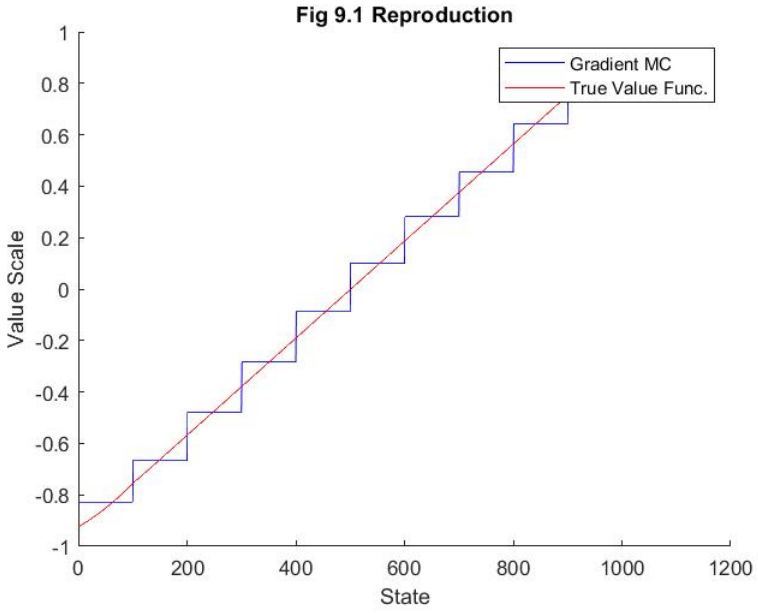
EEC-289A Reinforcement Learning
Homework #6

Jonathan Dorsey
https://github.com/JonnyD1117/EEC-289A-RL/tree/main/HW

May 20, 2021

Problem #1:

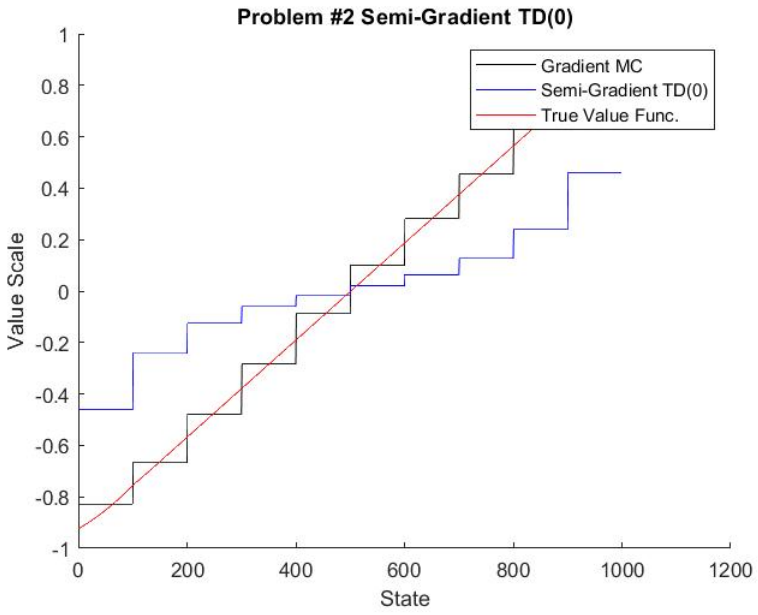
Reproduce Figure 9.1: This problem reproduces Figure 9.1 in Sutton & Barto, which illustrates a 1000-State Random Walk environment. The figure shows how the actual value function for the MDP (computed using Policy Evaluation), and the State Aggregation solution which uses a linear function to approximate the true value function of the environment. The approximate solution is obtained by using Gradient Monte Carlo Learning.



In this figure, the true value function is shown a mostly straight line, while the approximate value function is shown in a stair-step pattern. This pattern for the approximate solution is due to the state aggregation.

Problem #2:

In this problem, the same environment as in Problem #1 is evaluated using Semi-Gradient TD(0). However this solution uses the same settings and parameters as used in the previous problem. As such, the solution to this problem does not actually reproduce Figure 9.2 (Sutton & Barto), which uses the same algorithms on the same environment.



As can be seen, the Semi-Gradient TD(0) algorithm (using the same settings as Problem #1), the approximate value function obtained is biased and has not converged.

For Fun

In order to get the Semi-Gradient TD(0) algorithm to reprodce exactly like Figure 9.2, the algorithm requires more episodes of training. While a simple change, this has the effect of making the results from this algorithm reproduce Figure 9.2 far better than using the same settings as the first problem.

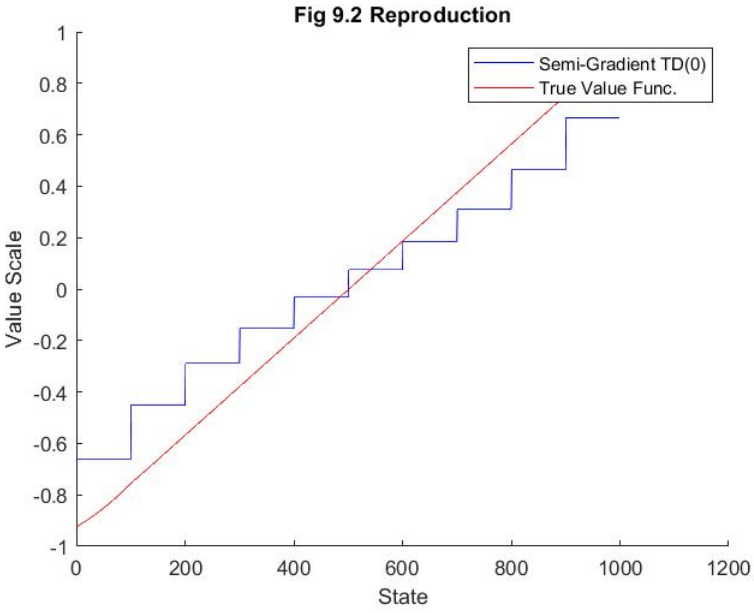


Fig 9.2 Reproduction
Semi-Gradient TD(0)
True Value Func.
Value Scale
State
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
200
400
600
800
1000
1200