

## Homework #2

# TD( $\lambda$ )

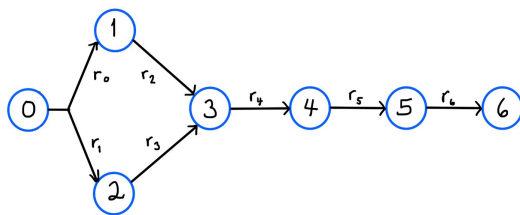
## Problem

### Description

Recall that the TD( $\lambda$ ) estimator for an MDP can be thought of as a weighted combination of the k-step estimators  $E_k$  for  $k \geq 1$ .

$$\text{TD}(\lambda) = \sum_{k=1}^{\infty} (1 - \lambda) \lambda^{k-1} E_k.$$

Consider the MDP described by the following state diagram. (Assume the discount factor is  $\gamma = 1$ .)



### Procedure

- Find a value of  $\lambda$ , strictly less than 1, such that the TD estimate for  $\lambda$  equals that of the TD(1) estimate. Round your answer for  $\lambda$  to three decimal places.
- This HW is designed to help solidify your understanding of the Temporal Difference algorithms and k-step estimators. You will be given the probability to State 1 and a vector of rewards  $\{r_0, r_1, r_2, r_3, r_4, r_5, r_6\}$
- You will be given 10 test cases for which you will return the best lambda value for each. Your answer must be correct to 3 decimal places. You may use any programming language and libraries you wish.



## Examples

The following examples can be used to verify your calculation is correct.

- Input: probToState=0.81, valueEstimates={0.0,4.0,25.7,0.0,20.1,12.2,0.0}, rewards={7.9,-5.1,2.5,-7.2,9.0,0.0,1.6}, Output: 0.6226326309908364
- Input: probToState=0.22, valueEstimates={0.0,-5.2,0.0,25.4,10.6,9.2,12.3}, rewards={-2.4,0.8,4.0,2.5,8.6,-6.4,6.1}, Output: 0.49567093118984556
- Input: probToState=0.64, valueEstimates={0.0,4.9,7.8,-2.3,25.5,-10.2,-6.5}, rewards={-2.4,9.6,-7.8,0.1,3.4,-2.1,7.9}, Output: 0.20550275877409016

## Resources

The concepts explored in this homework are covered by:

- Lectures
  - Lesson 3: TD and Friends
- Readings
  - Sutton (1988)

## Submission Details

**The due date is indicated on the Canvas page for this assignment.**

Make sure you have set your timezone in Canvas to ensure the deadline is accurate.

To complete the assignment calculate answers to the specific problems given and submit results at <https://rldm.herokuapp.com>

