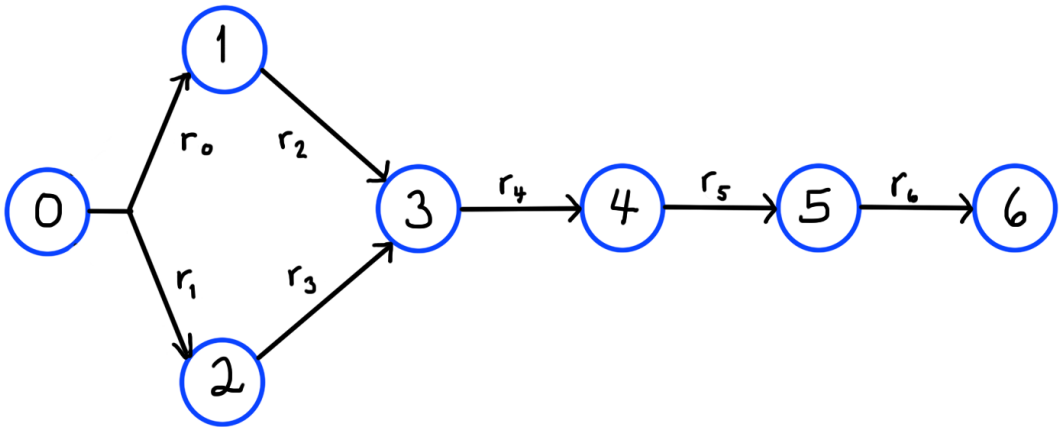
Homework 2

Problem

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



Find a value of λ, strictly less than 1, such that the TD estimate for λ equals that of the $TD(1)$ estimate. Round your answer for λ to three decimal places.

Here are a few examples to start with:

- **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