

Exercise: For the toy example (5 marks)

1) How many policies are possible?

enumerate the policies

2) For each of the policies

calculate  $V^\pi$

3) Use Value Iteration to

calculate  $V^*$

4) Use policy iteration to

improve the policy at each

step and eventually find  $V^*$

Create a Colab, send the link to me

~~Roll~~ - assign-1.ipynb

Finite horizon we used  $V_{T-1}^*$  to

find  $V_T^*$