Homework 3
Stats 232-C

Parth Shettiwar `parthshettiwar@g.ucla.edu`

February 19, 2022

1 Regarding code

The code generates all of the following 18 plots on the run. The names with which they are saved is intuitive. All the plots are saved in the same folder as the code.

2 Value Table Visualisations

The trap states get negative values and highest value is attached to final goal as expected.

Value Table of Environment 1 and Trajectory to Goal A

	0	1	2	3	4	5	6
5	62.613	66.961	71.538	76.356	81.427	86.765	87.179
4	62.933	67.298	71.893	76.729	81.821	87.179	92.821
3	62.56	66.905	71.479	-22.271	81.427	86.765	87.179
2	62.166	66.491	66.905	-22.644	81.013	81.427	81.821
1	61.752	62.166	62.56	-23.038	75.962	76.356	76.729
0	57.664	58.058	58.432	-27.836	71.164	71.538	71.893

Value Table of Environment 2 and Trajectory to Goal A

	0	1	2	3	4	5	6
5	62.613	66.961	71.538	76.356	81.427	86.765	87.179
4	62.933	67.298	71.893	76.729	81.821	87.179	92.821
3	62.56	66.905	71.479	-22.271	81.427	86.765	87.179
2	62.166	66.491	70.356	-22.644	81.013	81.427	81.821
1	61.92	66.232	70.771	75.548	75.962	76.356	76.729
0	61.547	65.839	70.356	-27.836	71.164	71.538	71.893

Value Table of Environment 1 and Trajectory to Goal B

	0	1	2	3	4	5	6
5	61.547	65.839	70.356	70.771	71.164	71.538	71.893
4	61.92	66.232	70.771	75.548	75.962	76.356	76.729
3	61.547	65.839	70.356	-22.644	81.013	81.427	81.821
2	61.153	65.424	65.839	-22.271	81.427	86.765	87.179
1	60.739	61.153	61.547	-22.271	81.821	87.179	92.821
0	56.702	57.095	57.469	-22.271	81.427	86.765	87.179

Value Table of Environment 2 and Trajectory to Goal B

	0	1	2	3	4	5	6
5	61.547	65.839	70.356	70.771	71.164	71.538	71.893
4	61.92	66.232	70.771	75.548	75.962	76.356	76.729
3	62.166	66.491	70.356	-22.644	81.013	81.427	81.821
2	62.56	66.905	71.479	-22.271	81.427	86.765	87.179
1	62.933	67.298	71.893	76.729	81.821	87.179	92.821
0	62.56	66.905	71.479	-22.271	81.427	86.765	87.179

Figure 1: Value Tables for all environments and trajectories

3 Policy Table Visualisations

As expected, all arrows mostly point towards the final goal state, avoiding the trap states.
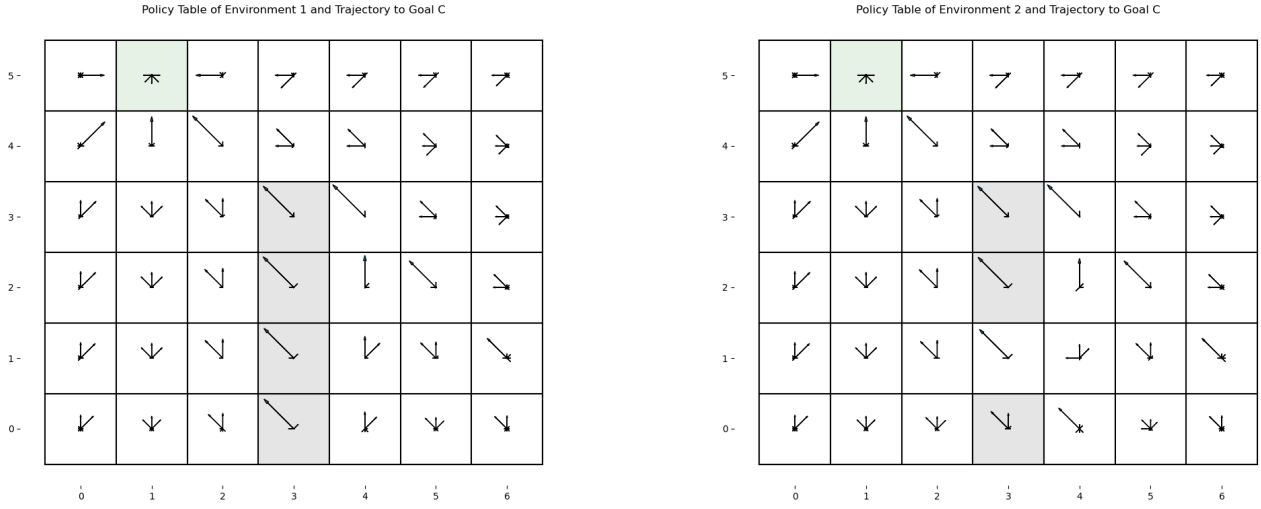
Figure 2: Policy Tables for all environments and trajectories

4 Graph Plots for Posterior Probability

We observe that after 3-4 time stamps, always the policy which have been computed for a particular goal gets highest posterior probability and other 2 policies die down. This is obvious since the way policies have been compute always favourite their own goal trajectory.
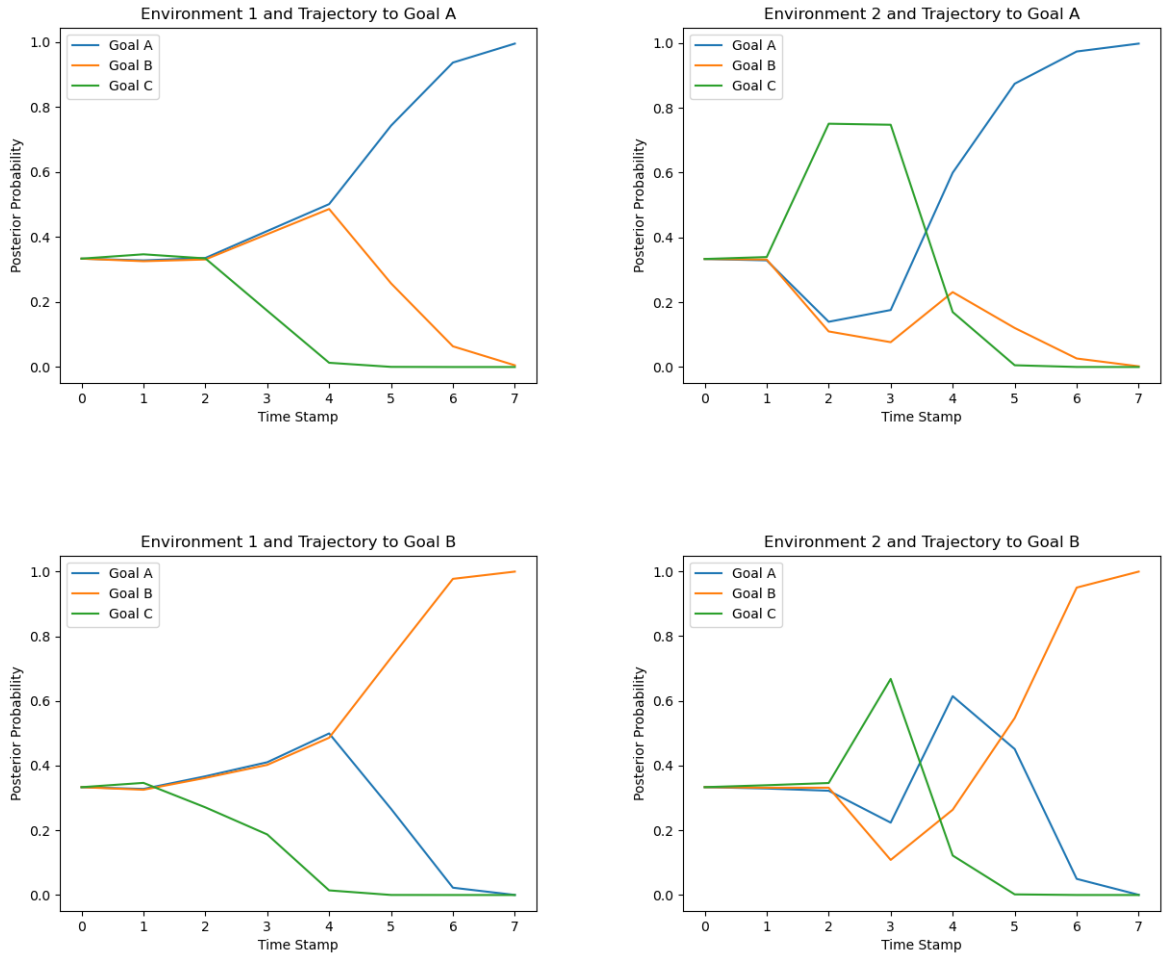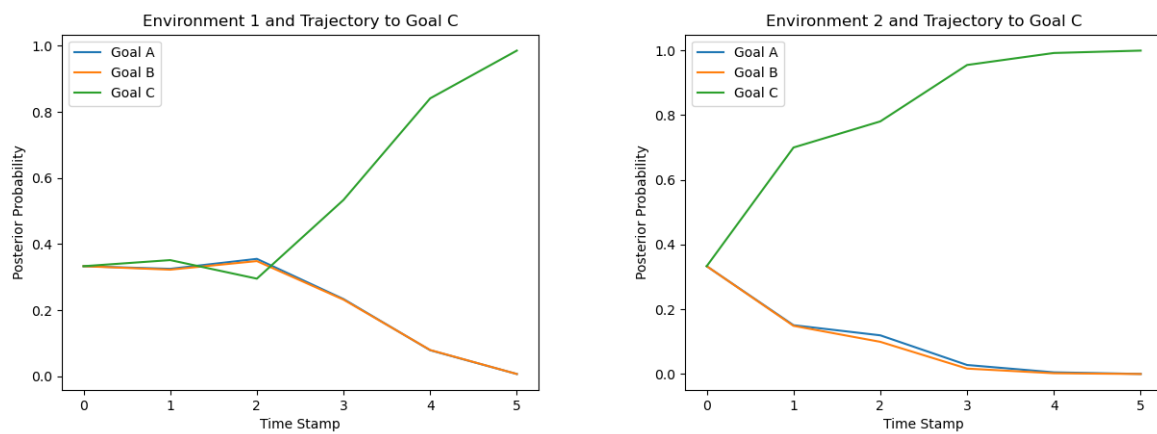
Figure 3: Posterior Probability graphs for all environments and trajectories