1 Problem 1

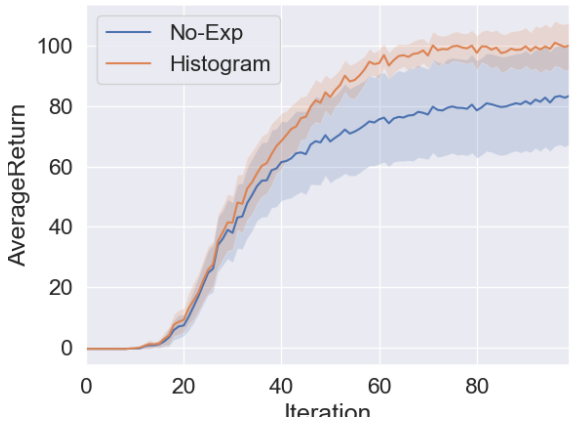

Figure 1: Problem 1

Please see Figure 1.

2 Problem 2

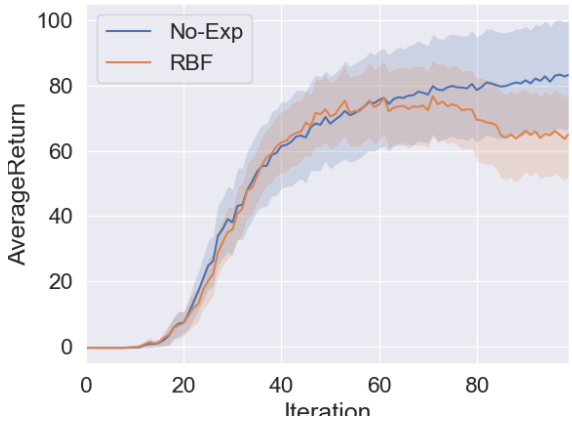

Figure 2: Problem 2

Please see Figure 2.

3 Problem 3

Please see Figure 3.

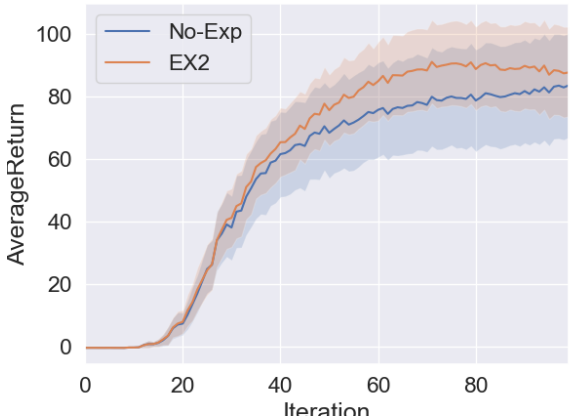

Figure 3: Problem 3

4 Problem 4

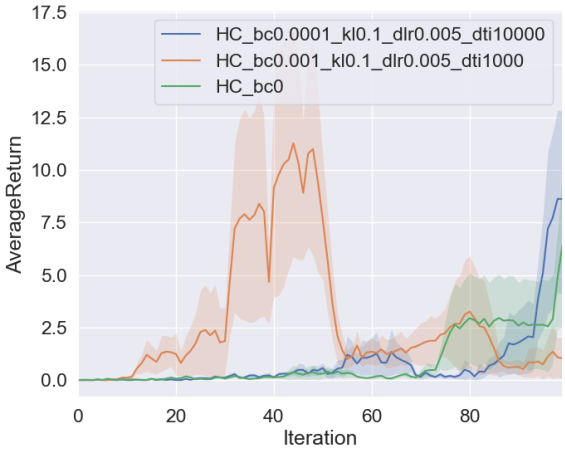

Figure 4: Problem 4

Please see Figure 4. I guess with larger bonus weight the agent quickly find the best trajectory, but since it will keep visiting the optimal state, the modified reward for suboptimal states will be larger, and the behavior becomes suboptimal. With appropriate weight it learns slower but is more stable.