1 Problem 1

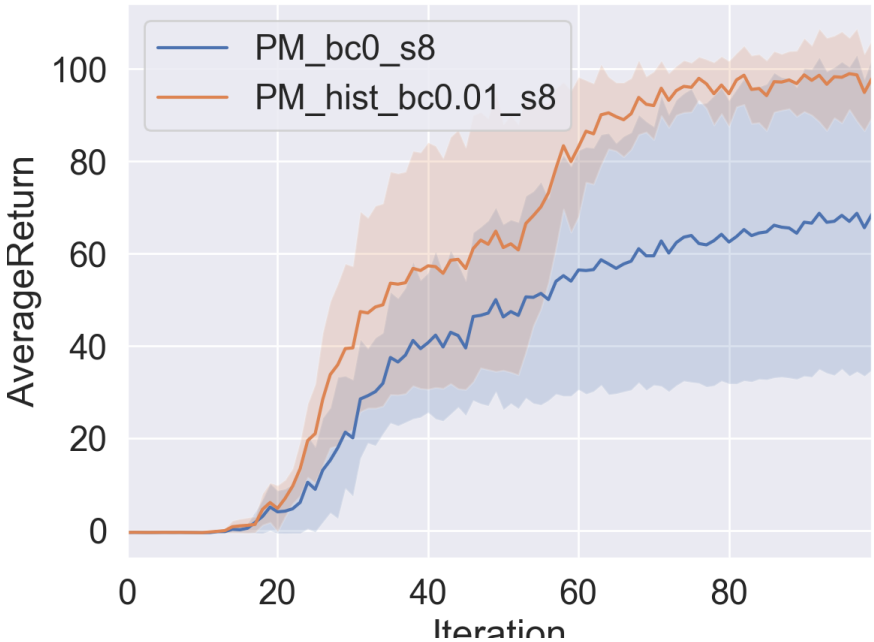

Figure 1: Comparison for agent with histogram-based exploration and agent with no exploration for PointMass.

2 Problem 2

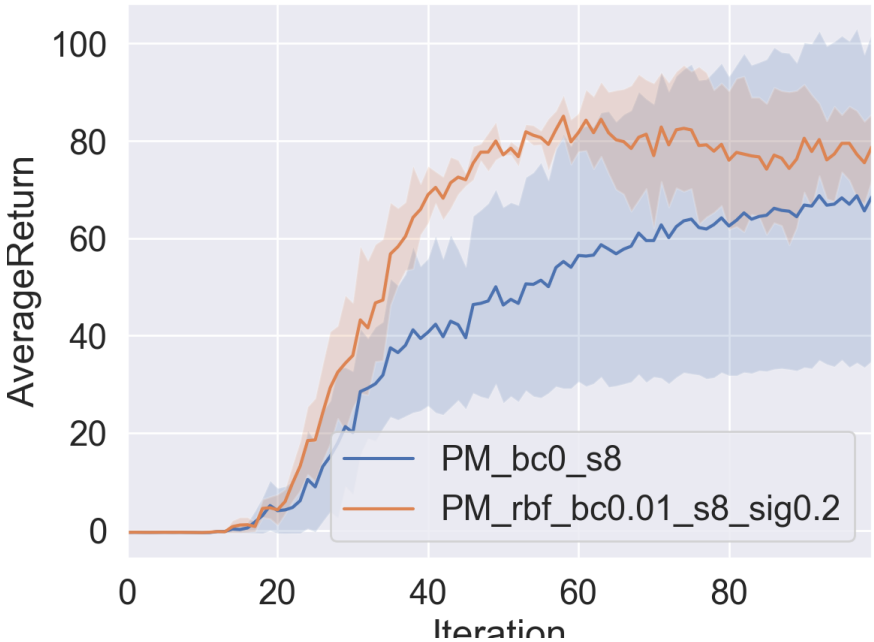

Figure 2: Comparison for agent with KDE-based exploration and agent with no exploration for PointMass.

3 Problem 3

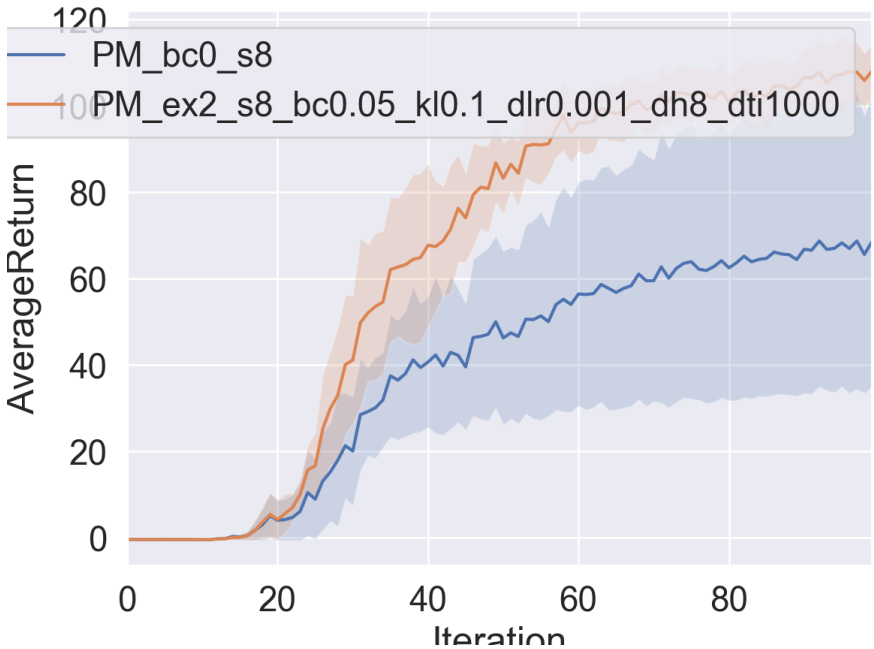

Figure 3: Comparison for agent with EX2-based exploration and agent with no exploration for PointMass.

4 Problem 4

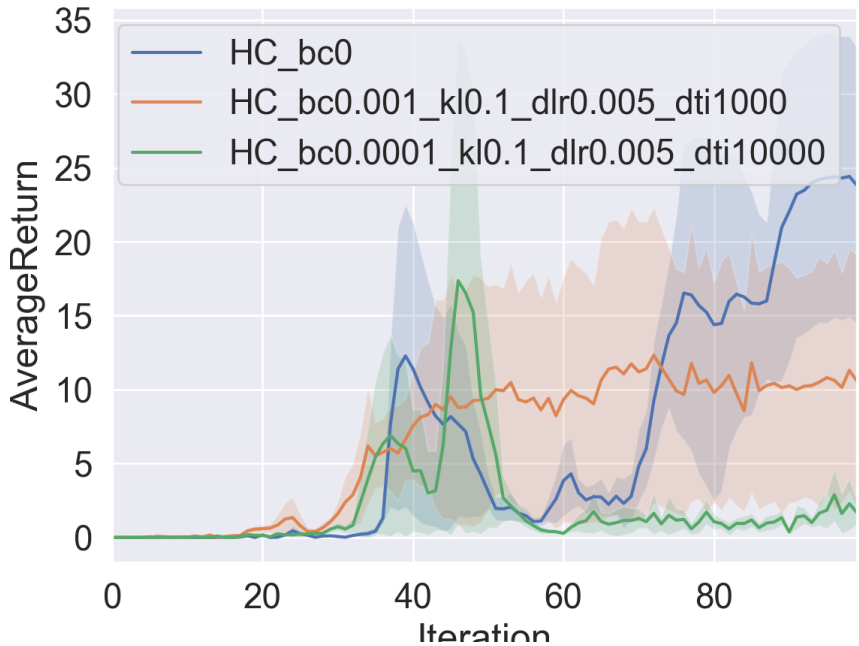

Figure 4: Comparison for two agents with EX2-based exploration and another agent with no exploration for HalfCheetah.

(1)The learning curve is going up and down is because that even after the agents have already reached the optimal states, they are still incentivized by the reward bonus to explore somewhere else. Thus the agents wouldn't stop exploring, which would result in suboptimal returns.
(2)Different bonus parameter will affect agent's performance. A bigger bonus parameter will keep driving the agent to explore somewhere else and have a different learning curve.