CD294-112 Deep Reinforcement Learning HW5a: Exploration

Ke Wang, EECS Ph.D. student, kewang@berkeley.com

2018/10/7

1 Problem 1: Discrete States Exploration

Here, we present a plot with 2 curves comparing an agent with histogram-based exploration and an agent with no exploration.

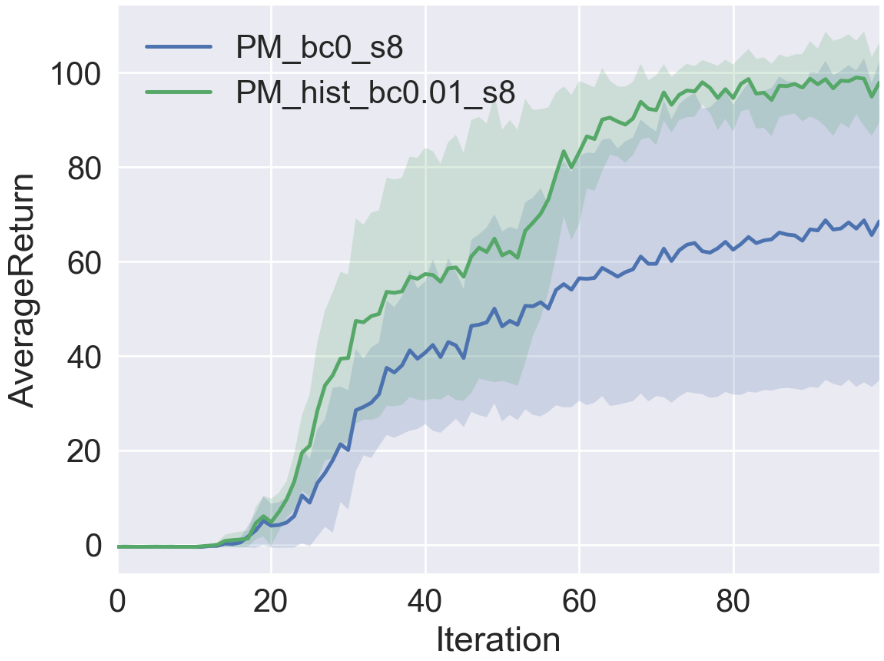

Figure 1: Average return of agent with histogram-based exploration and with no exploration

.

As shown in the figure, the average return of histogram-based exploration is higher than the one without exploration.

2 Problem 2: Continuous States Exploration

Here, we present a plot with 2 curves comparing an agent with KDE-based exploration and an agent with no exploration for PointMass.

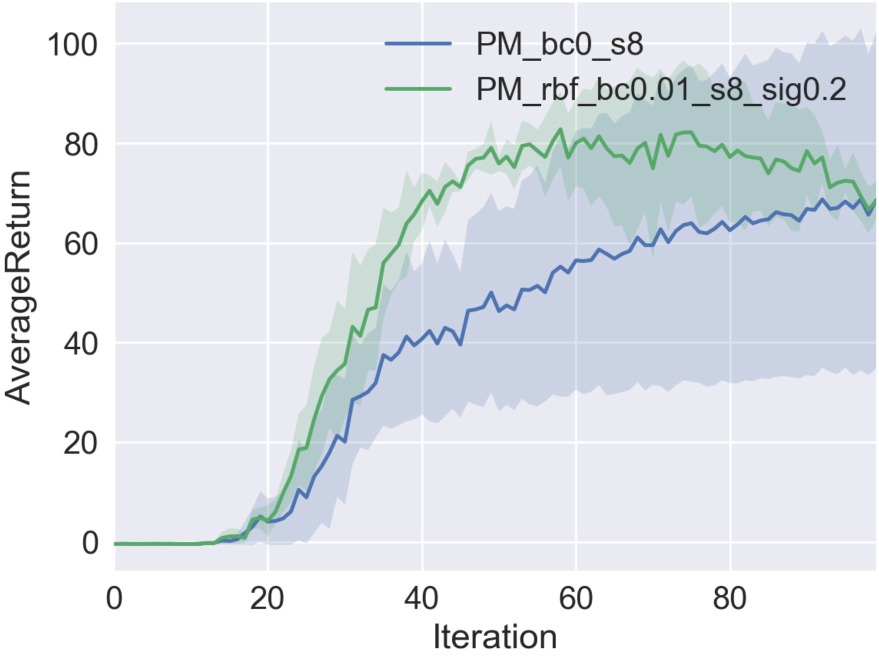

Figure 2: Average return of agent with KDE-based exploration and with no exploration
.

As shown in Figure 2, KDE-based exploration can have a higher average return with small variance.

3 Problem 3: Continuous States Exploration EX2 discriminator

Here, we present a plot with 2 curves comparing an agent with EX2-based exploration and an agent with no exploration for PointMass.

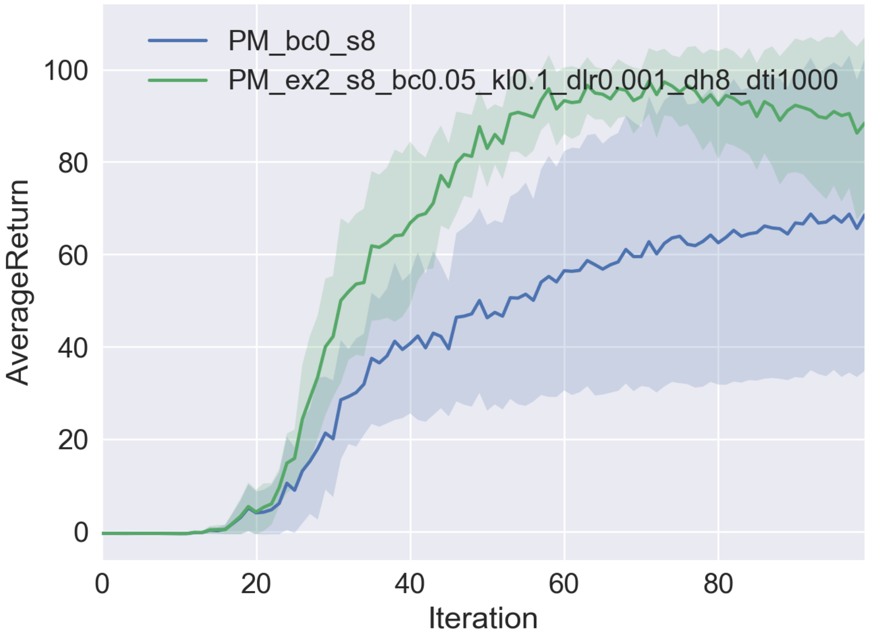

Figure 3: Average return of agent with EX2-based exploration and with no exploration

.

Using EX2-based exploration can result in higher average return with small variance.

4 Problem 4: Continuous States Exploration EX2 discriminator for HalfCheetah

Here, we present a plot with 3 curves comparing an agent with EX2-based exploration and an agent with no exploration for HalfCheetah.

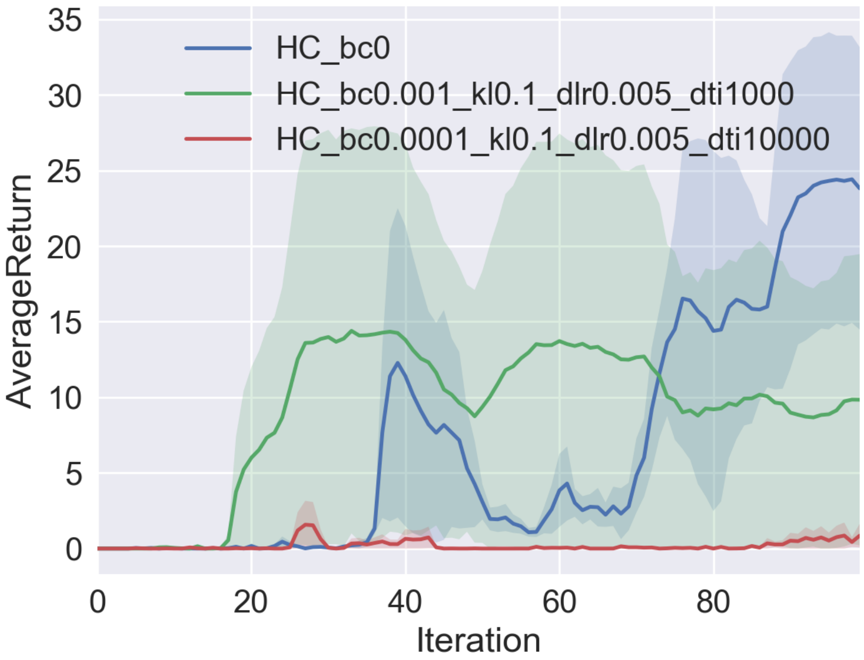

Figure 4: Average return of agent with EX2-based exploration and with no exploration for HalfCheetah .

Short Answer:For HalfCheetah task, using EX2-based learning curve gets a higher average reward when the iteration times is low but lower average reward when the iteration times is high. Average return for $\alpha = 0.0001$ is lower than $\alpha = 0.001$.

More details are described in the code.