

CS294-112 Deep Reinforcement Learning HW4: Model-Based RL

Louis TILLOY

Problem 1

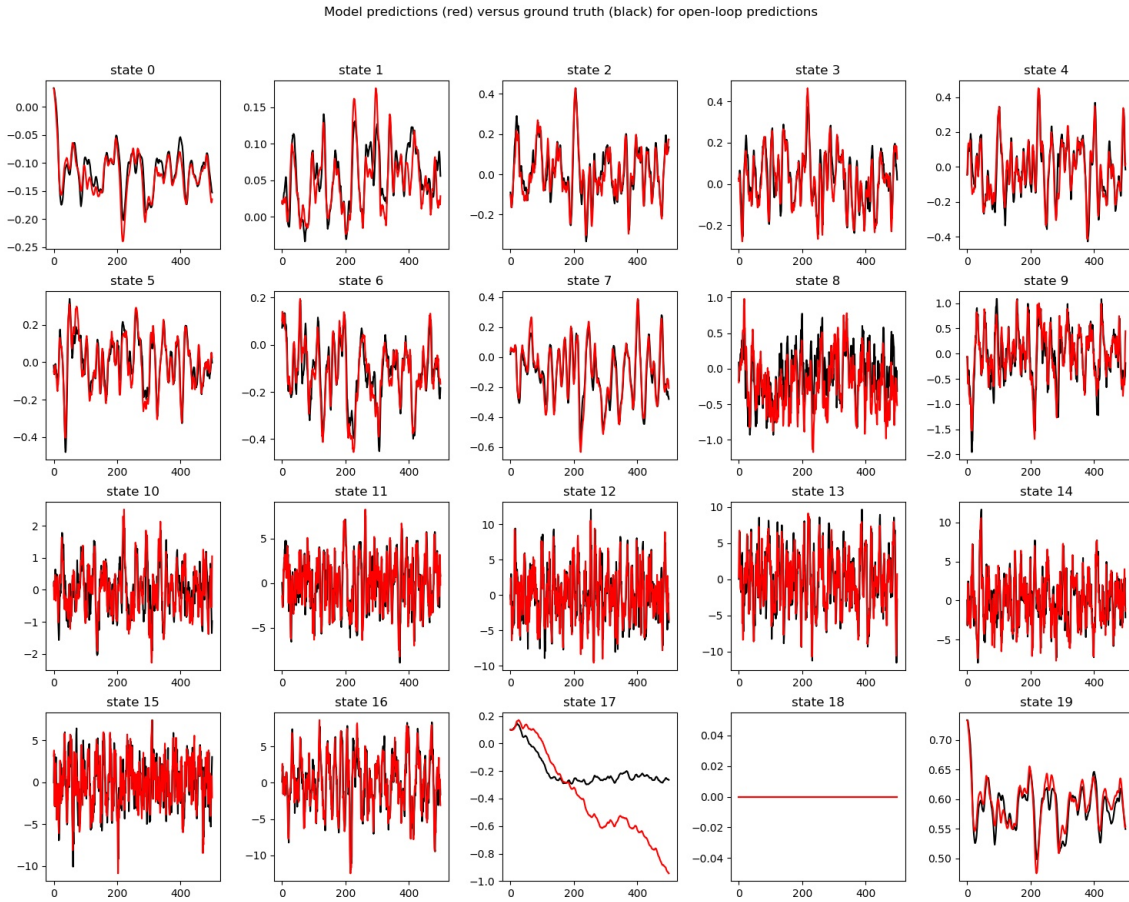


FIGURE 1 – dynamics model predictions

We can see that the predictions are the most inaccurate for the dimension 17. A difference between the dimension 17 and the others dimensions is that the values have a global tendency to decrease. The neural network learned to match this average decrease with some error that is propagating at each time step, therefore creating a bias roughly proportional with the number of steps.

Problem 2

The ReturnAvg and ReturnStd for the random policy and for the model-based controller trained on the randomly gathered data are the following :

random policy

$$\begin{aligned} \textit{ReturnAvg} &= -132.072 \\ \textit{ReturnStd} &= 21.5319 \end{aligned}$$

model-based controller

$$\begin{aligned} \textit{ReturnAvg} &= 16.2403 \\ \textit{ReturnStd} &= 25.1196 \end{aligned}$$

Problem 3a

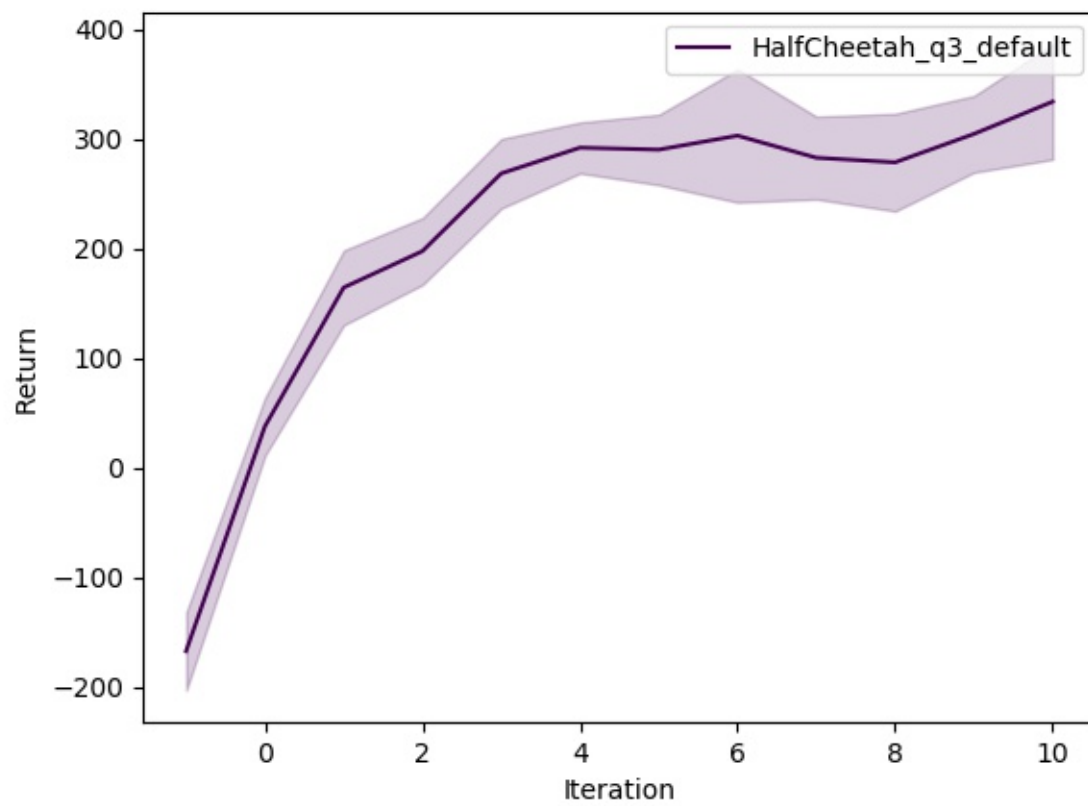


FIGURE 2 – Performance evolution of the model-based reinforcement learning with default parameters

Problem 3b

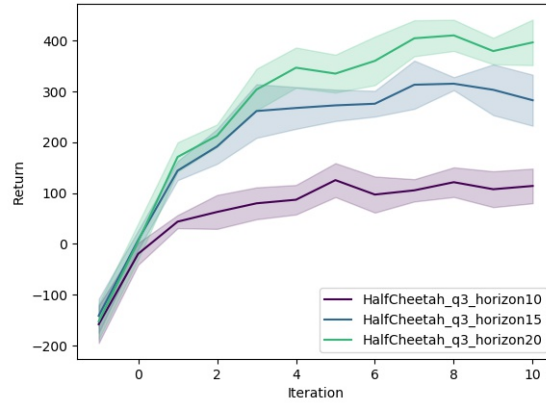


FIGURE 3 – Plot comparing performance when varying the MPC horizon

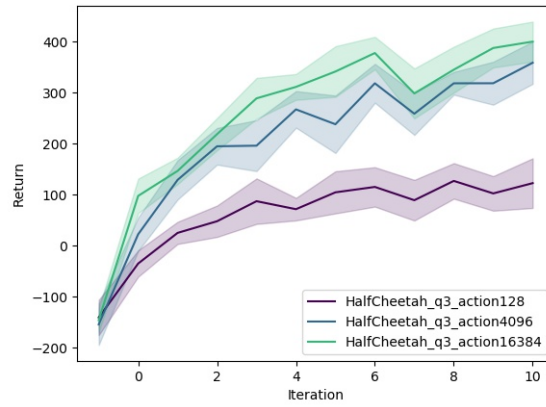


FIGURE 4 – Plot comparing performance when varying the number of randomly sampled action sequences used for planning

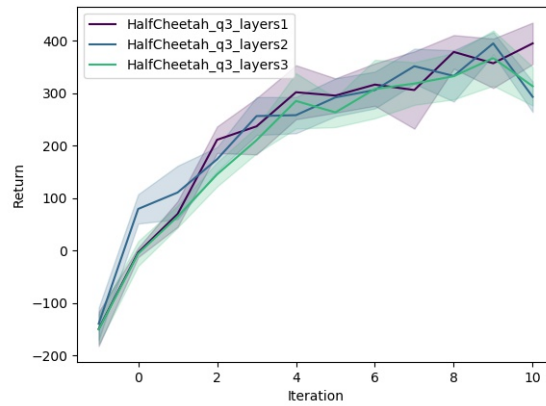


FIGURE 5 – Plot comparing performance when varying the number of neural network layers for the learned dynamics model