CS294-112 Assignment 4: Model-Based RL

Fan Zhang

Problem 1: Dynamics Model

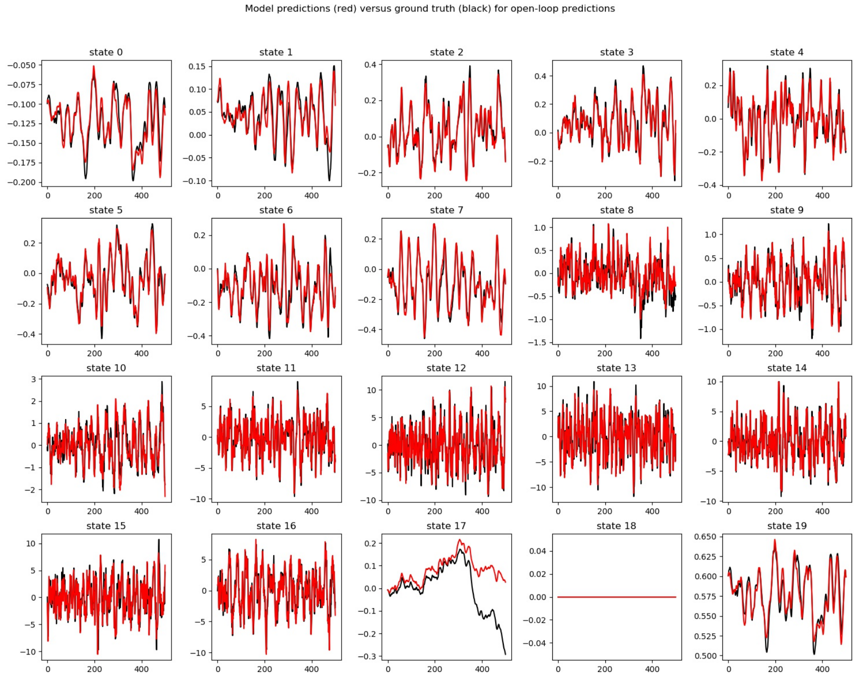


Figure 1: Trained dynamics model(red) versus ground truth(black) for open-loop predictions

Explanation: From the figure that the state 17 is the most inaccurate one among all 20 dimensions of states. The state 17 has a trend to increase or decrease so the error could be accumulated through the trajectory. If there is an error during a time stamp and then for the rest of trajectory this error will become larger. But for other states the value is more like random noises so the former error will not influence the later value or it won't be accumulated for a long time. The state 18 is another situation that it's value remains nearly the same so the prediction is easy and there is almost no error. It is like the long-term dependency and short-term dependency problem. For random distributed state dimensions the current value relies little on previous value so the error won't be accumulated.

Problem 2: Action Selection

Policy	ReturnAvg	ReturnStd
Random	-142.3057948	25.56133356
Model-based(Test 1)	-18.98624155	27.53356139
Model-based(Test 2)	19.06106592	29.43963377
Model-based(Test 3)	-6.062103844	19.7160912
Model-based(Test 4)	1.90644462	27.93877017

Table 1: For random policy, the average return is about -150 and the average std is about 25.5, and for model-based policy, the average return changes with different tests. I tested four times and the average return fluctuated between -20 to 20 and the average of average return is around 0 while the average std is around 30.

Problem 3a: MBRL with on-policy data collection

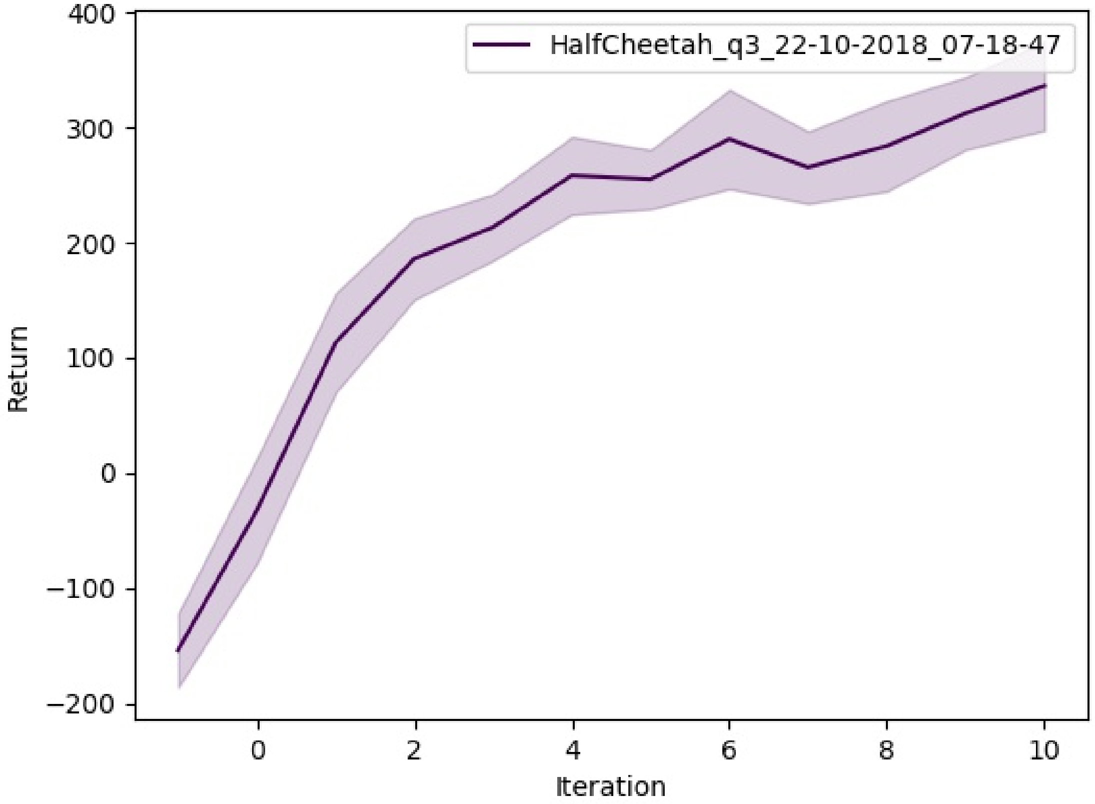
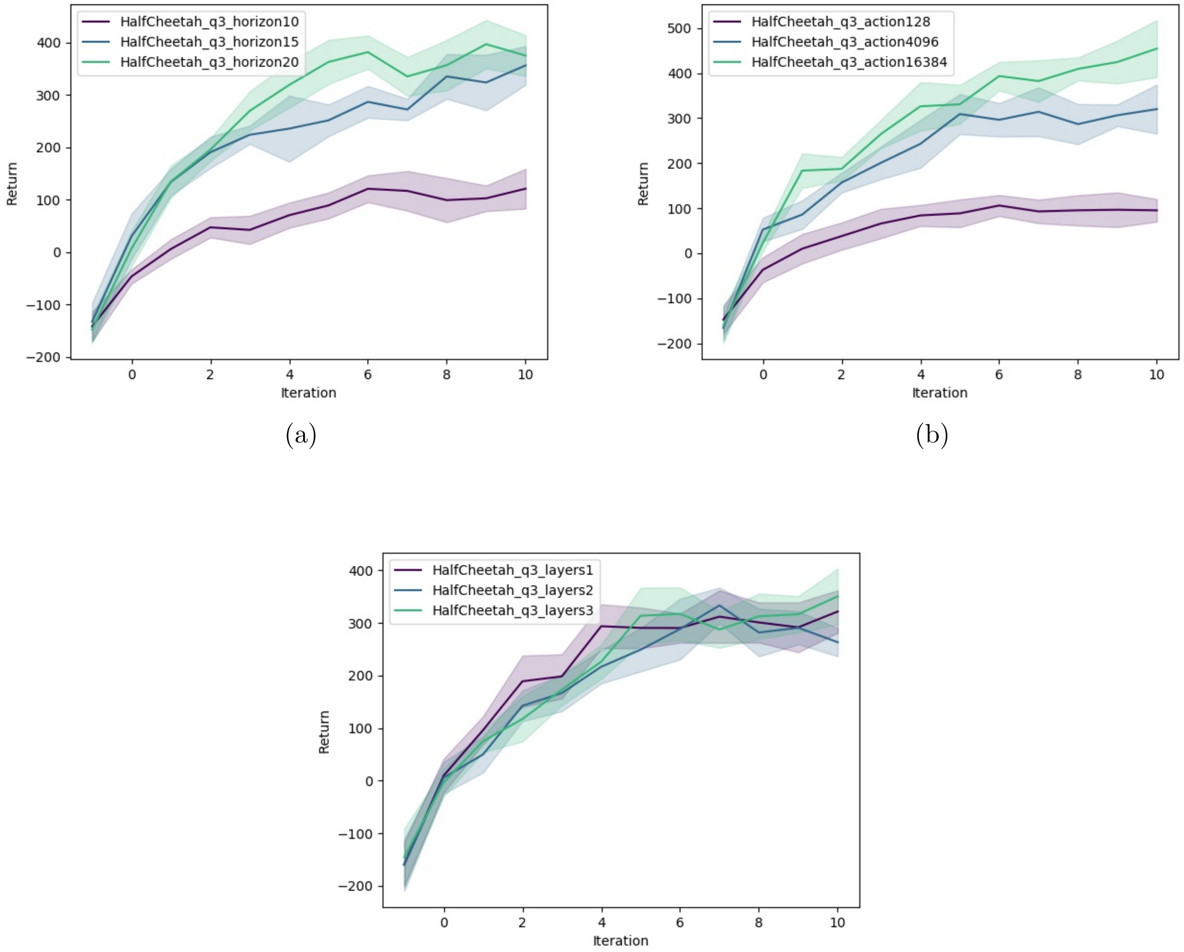


Figure 2: Return versus iteration with on-policy model-based RL, by the iteration 10, the return has arrived 300.

Problem 3b: Performance with hyperparameters



(a)

(b)

(c)

Figure 3: The performance of model-based RL when (a) Return versus iteration with different number of MPC horizons (b)Return versus iteration with different number of action sequences (c)Return versus iteration with different number of layers in dynamic model networks.