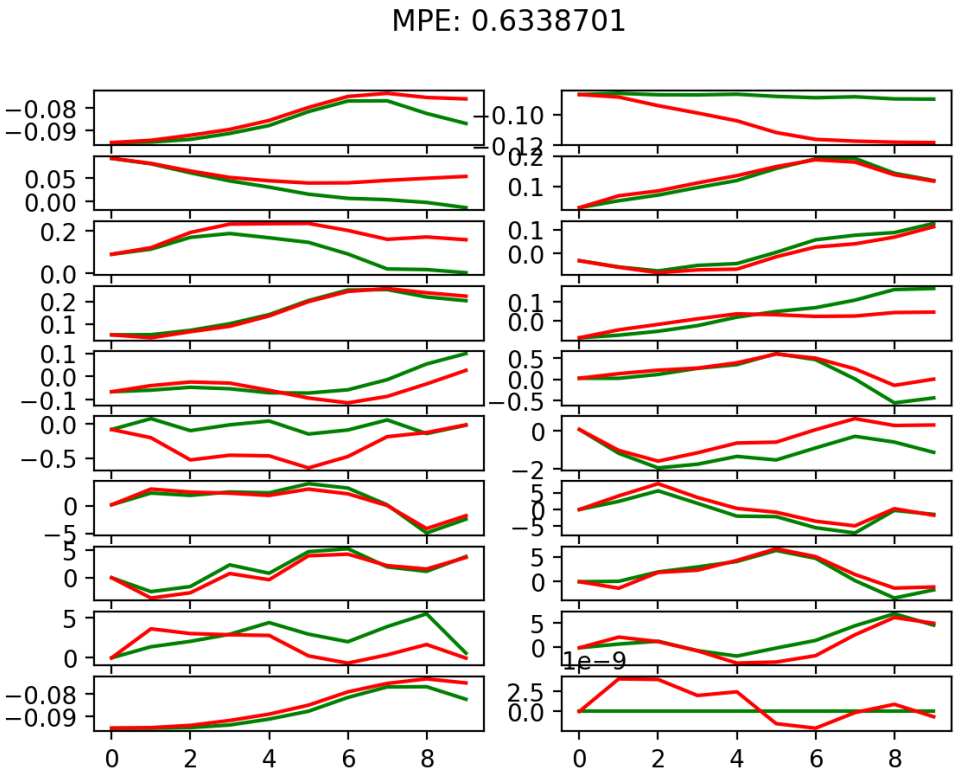
Assignment 4: Model Based RL

huseyinabanox@gmail.com

January 2023

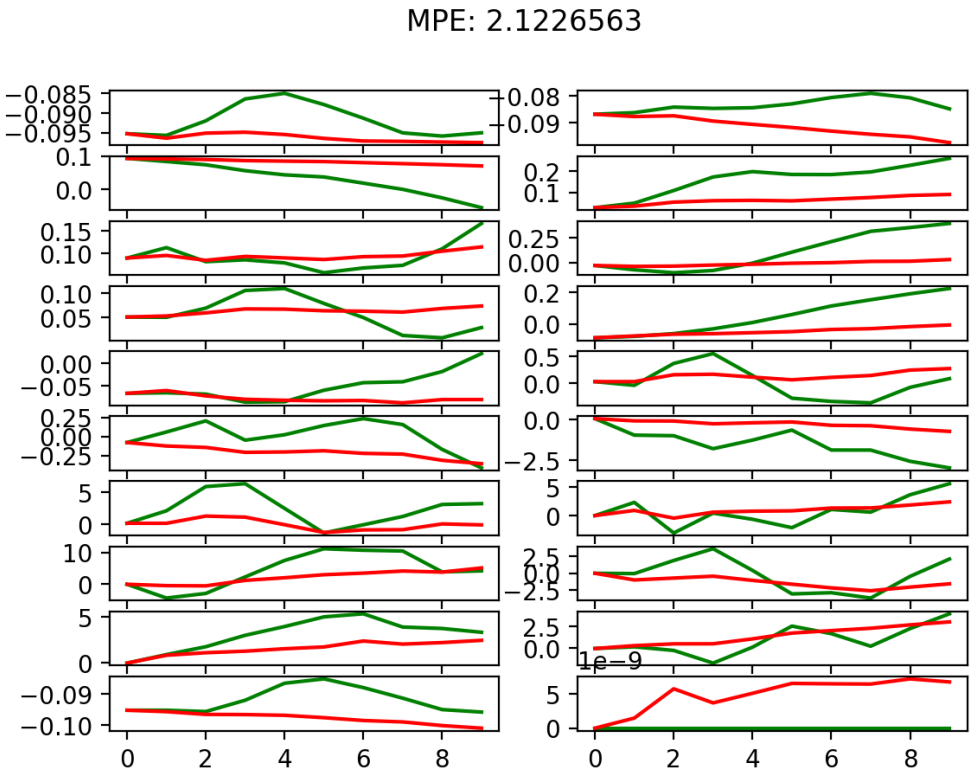
Problem 1

First Run



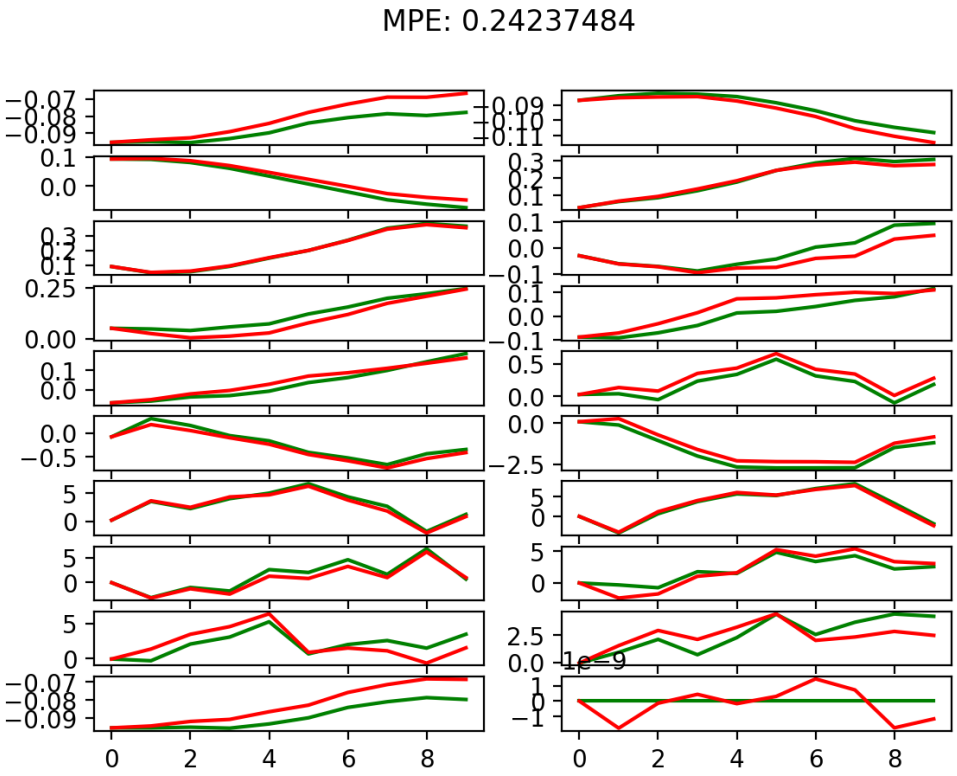
A small network is used. Results can be improved using a larger network.

Second Run



Small number of iterations are used. Results can be improved by increasing iteration count. MPE is the worst.

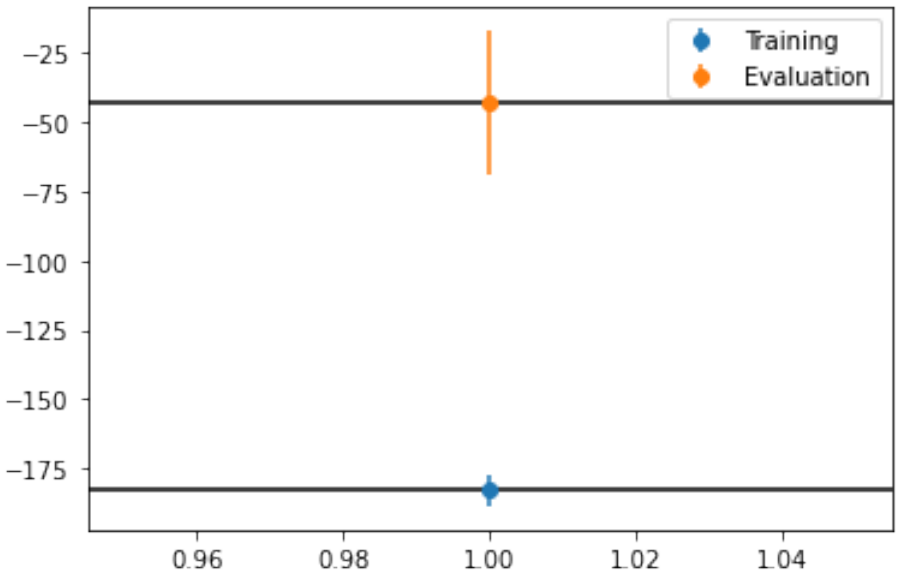
Third Run



Best results are obtained using a larger network and more iterations. MPE is the best.

Problem 2

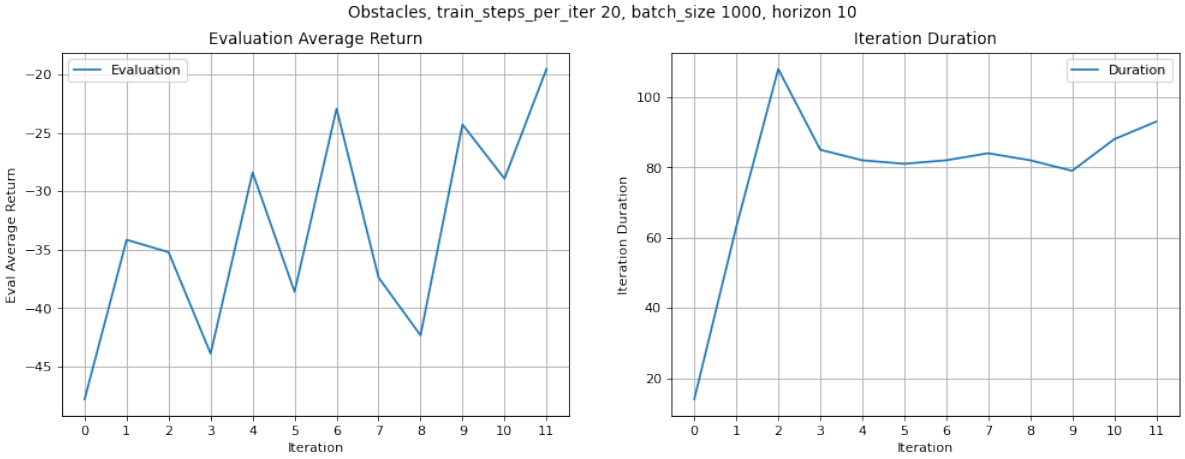
Train AverageReturn is expected to be around -160 and Eval AverageReturn is expected to be around -70 to -50.



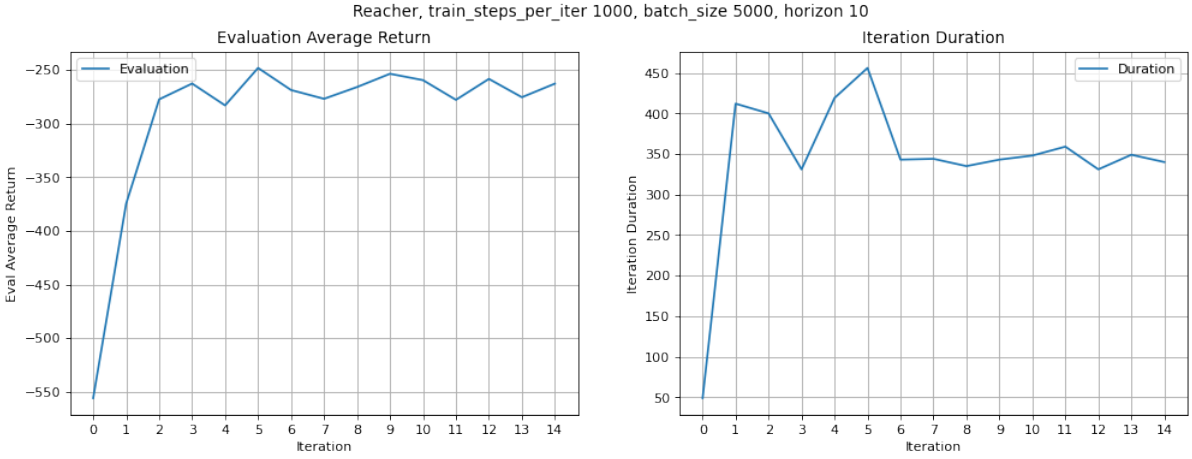
Actual returns are around the expected values.

Problem 3

MBRL algorithm with on-policy data collection and iterative model training. Rl trainer.py already aggregates your collected data into a replay buffer. Thus, iterative training means to just train on our growing replay buffer while collecting new data at each iteration using the most newly trained model.



Rewards of around -25 to -20 is expected for the obstacles env. The actual results are similar.

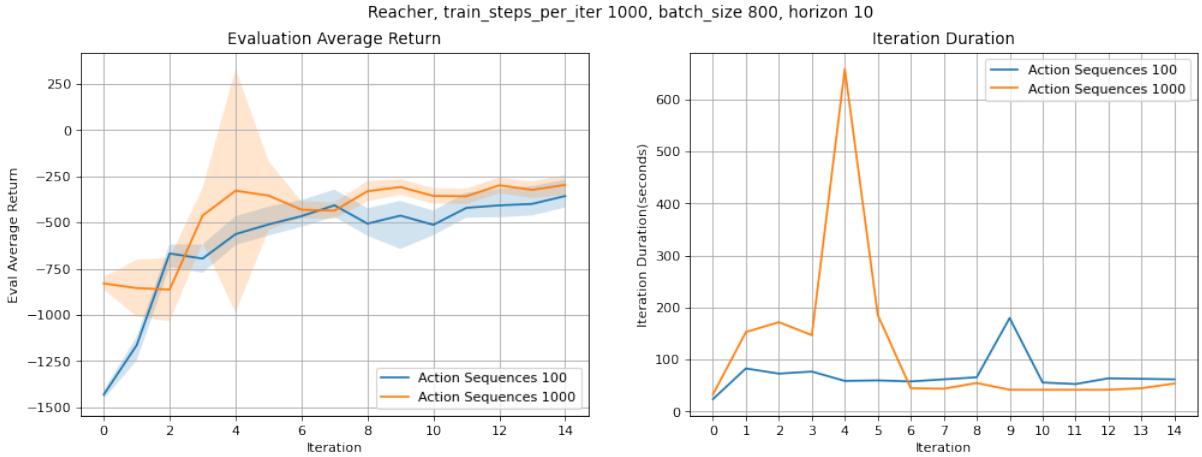


Rewards of around -250 to -300 is expected for the reacher env. The actual results are similar.

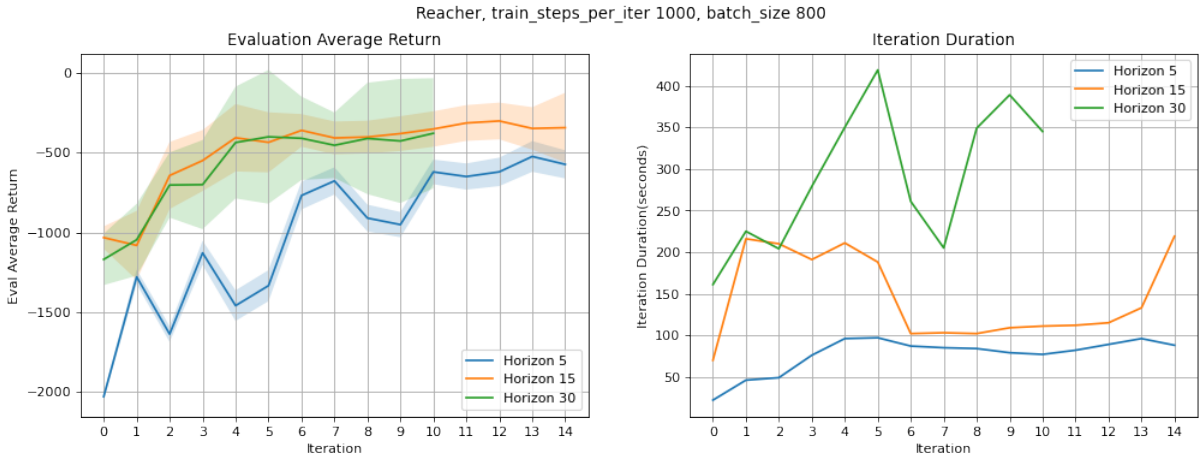
Rewards of around 250 to 350 is expected for the cheetah env. The actual results are similar.

Problem 4

The chart above shows effect of ensemble size. Increasing ensemble size improves average return but at the cost of training time. Intermediate values proposes the best trade off.

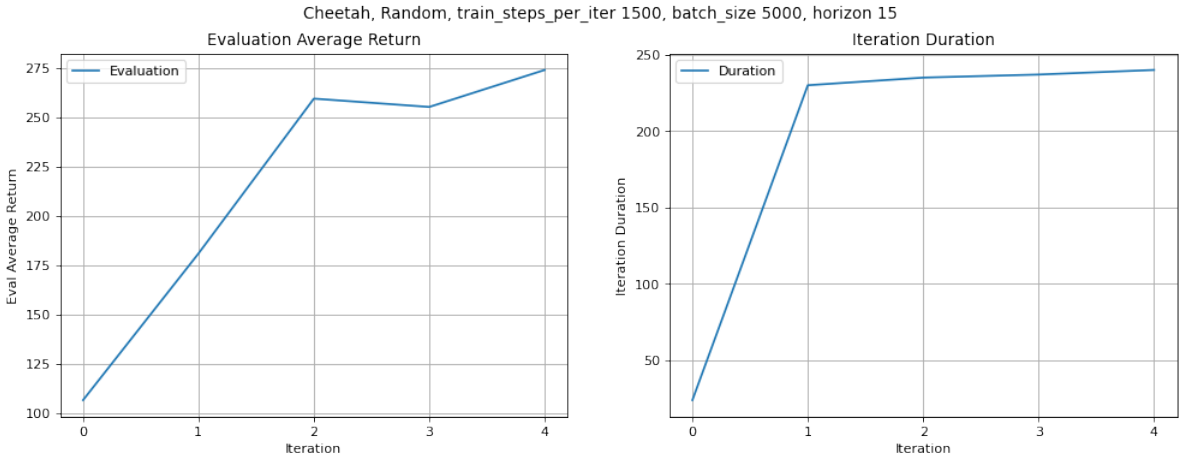


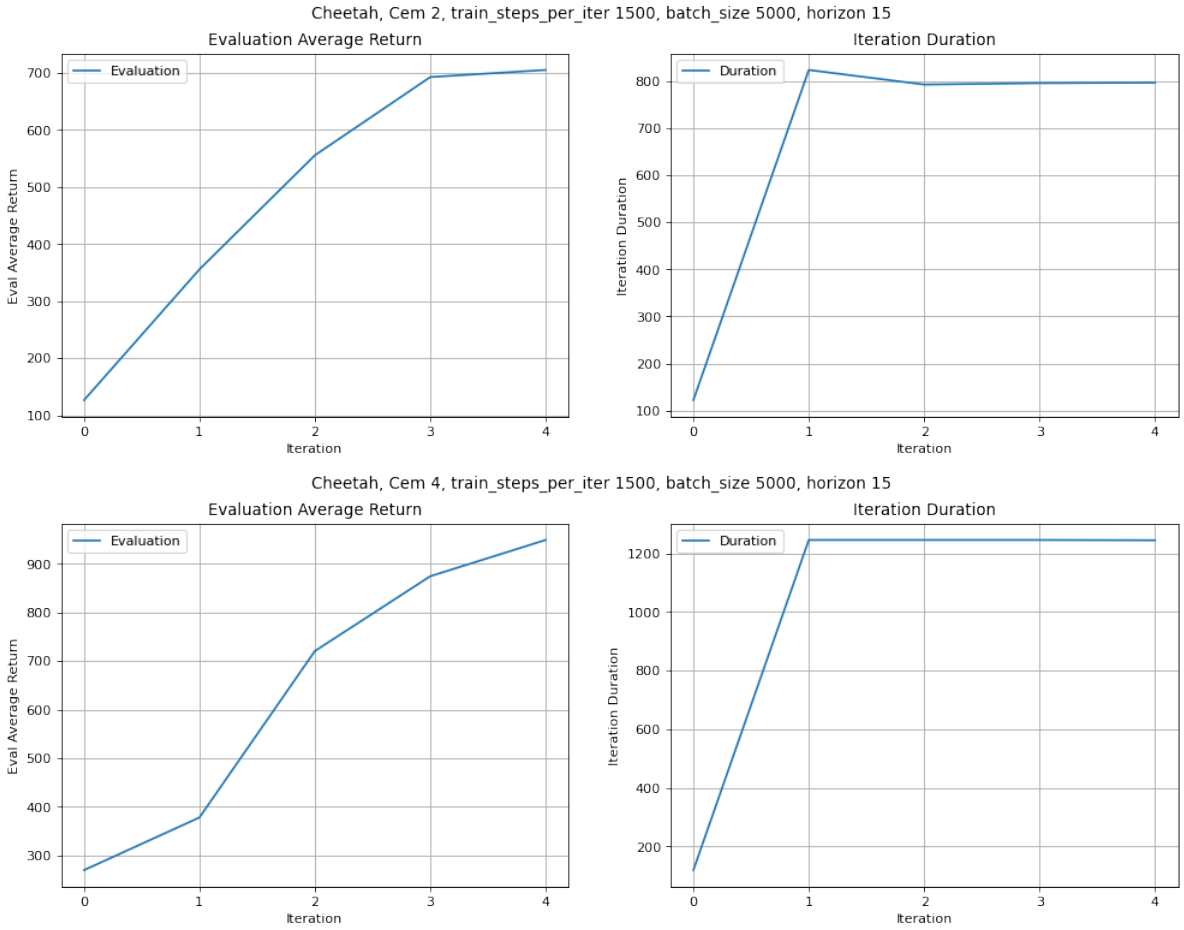
Increasing action sequence count improved average return slightly. Training time increases as well but the chart shows a mixed view. A more reliable duration chart should incorporate average of multiple runs.



Increasing action horizon improves average return but also increases variance. The final difference is not too much. The most apparent effect increasing the horizon is seems to be increasing the training speed. Again the best value is obtained using the intermediate values.

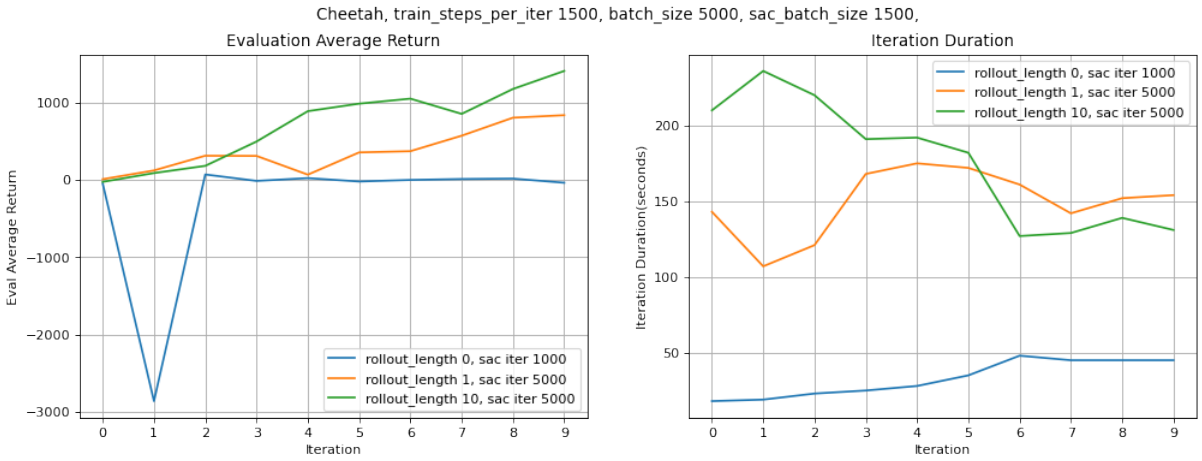
Problem 5





The charts show the effect of using cem. Using cem has a clear advantage over random sampling. Iteration duration increases similarly. 4 iteration cem yields better return at the cost of dramatically longer iterations. 2 iteration cem provides better value by providing a better return/duration ratio.

Problem 6



The chart shows the effect of collecting model trajectory while training sac policy. With 0 model trajectory (no model trajectory) training stagnates. With model trajectory of length 1, training improves. With model trajectory of length 10 learning is faster.