Assignment 4: Model-Based RL and Exploration

Andrew ID: `mukaiy`
NOTE: Please do **NOT** change the sizes of the answer blocks or plots.

1 Problem 1: Dynamics Model Training – [10 points total]

Theory questions

The third model performs the best, because it achieves the least $MPE = 0.07804489$.
More training steps per iteration improves convergence a lot, and larger MLP interpolates better.

Plot

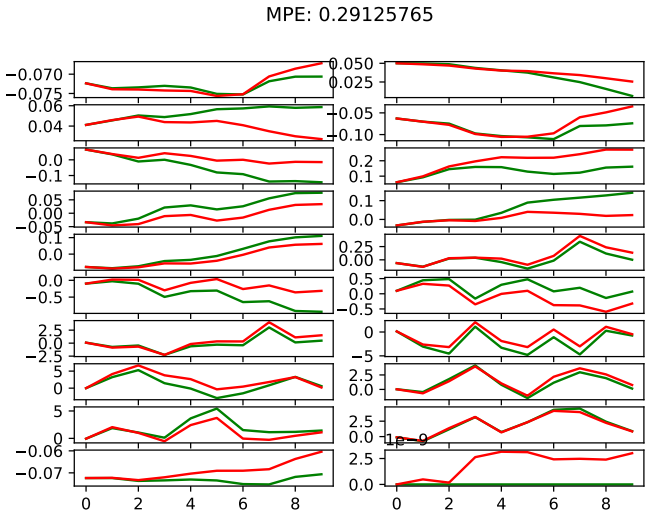


Figure 1: 500 training steps per iteration, 1 x 32 MLP.

Plot

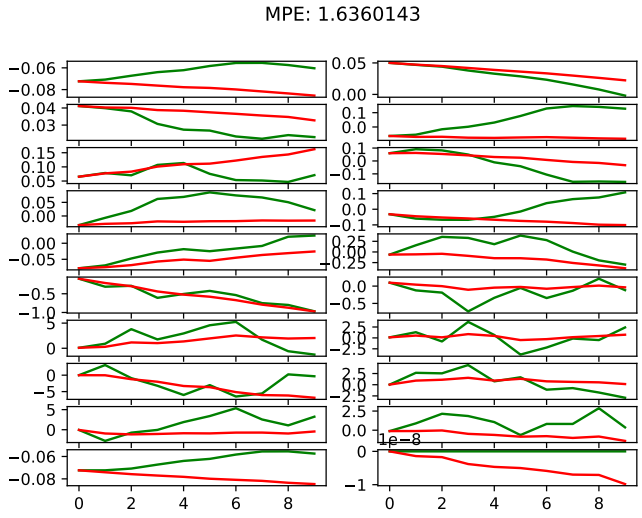


Figure 2: 5 training steps per iteration, 2 x 250 MLP.

Plot

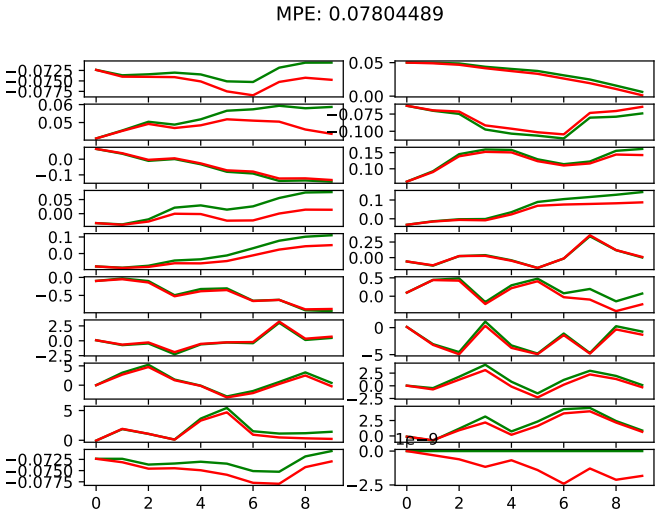


Figure 3: 500 training steps per iteration, 2 x 250 MLP.

2 Problem 2: Action Selection

Plot

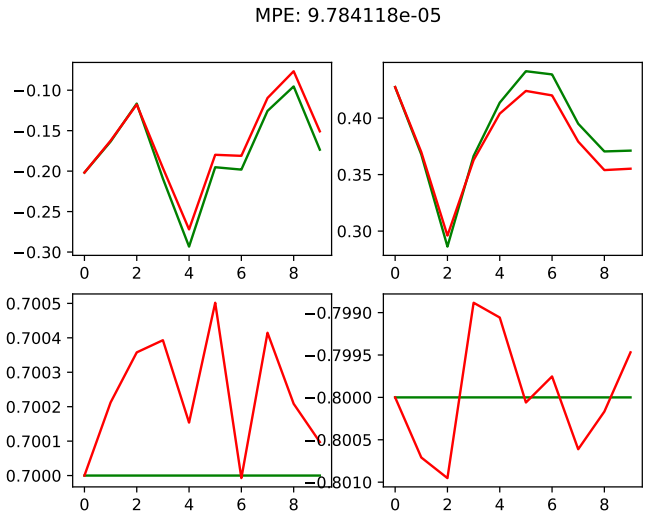


Figure 4: Optimizing the Dynamics Model where MPC needs to be used improves a lot.

3 Problem 3: Iterative Model Training

Plot

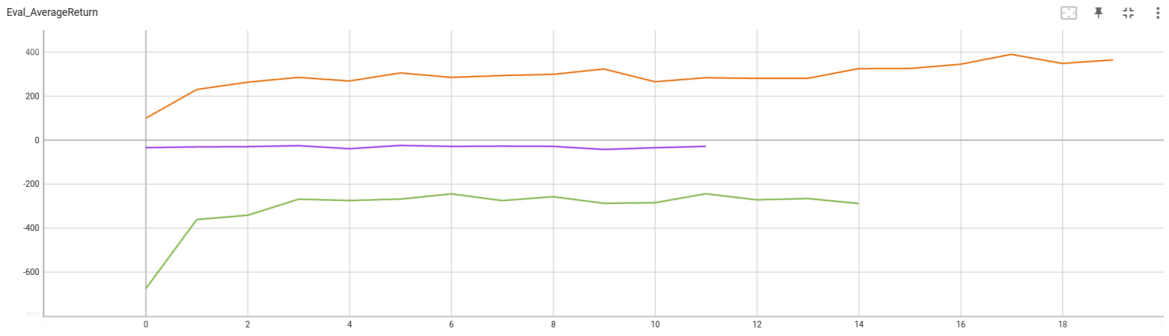


Figure 5: Average evaluation return over training iterations. Orange: cheetah. Purple: obstacles. Green: reacher.

Plot

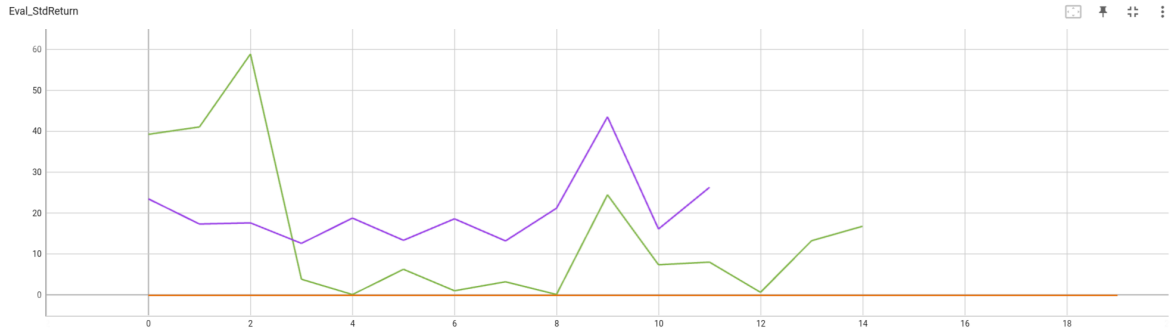


Figure 6: Standard deviation evaluation return over training iterations. Orange: cheetah. Purple: obstacles. Green: reacher.

Plot

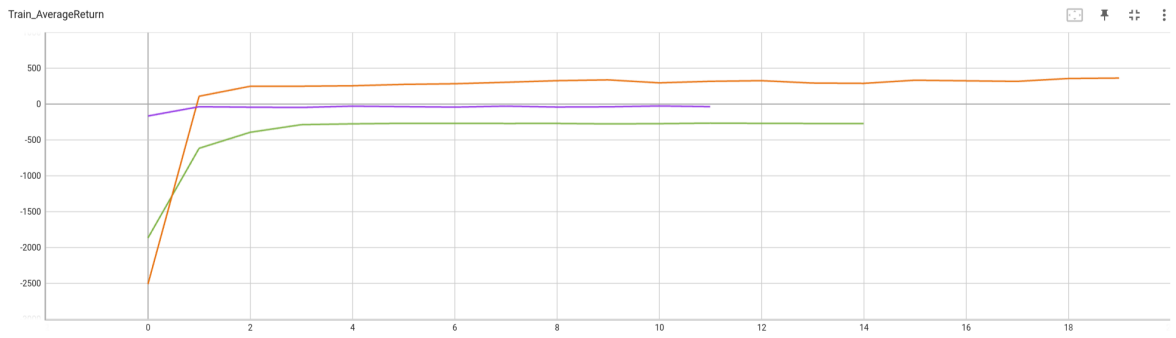


Figure 7: Average training return over training iterations. Orange: cheetah. Purple: obstacles. Green: reacher.

4 Problem 4: Hyper-parameter Comparison

Plot

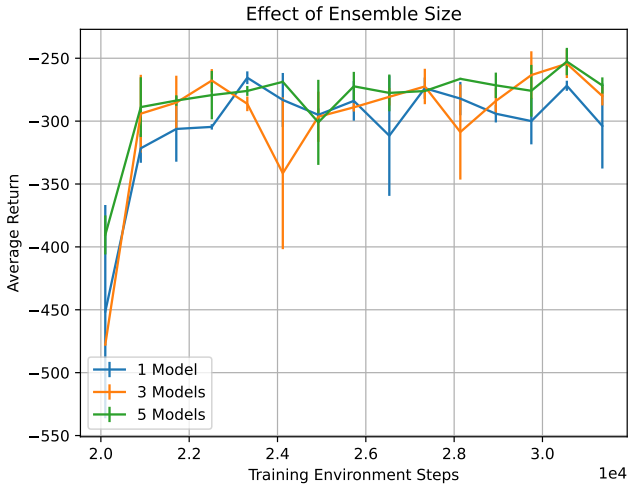


Figure 8: Effect of ensemble size seems not so significant, but it could be due to not big enough. It most likely only reduces variance partially.

Plot

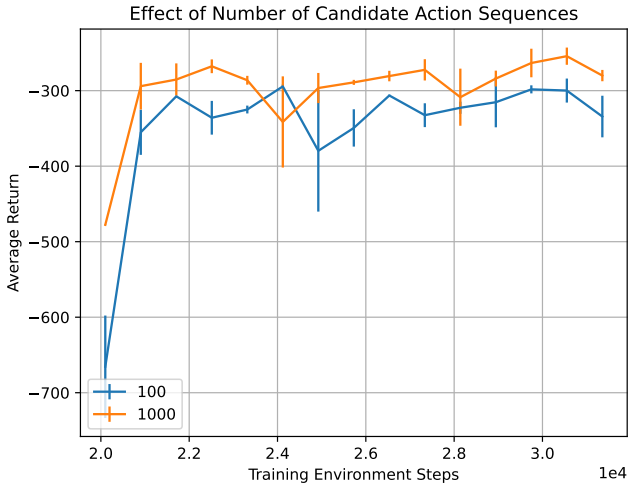


Figure 9: Effect of # candidate action sequences is more significant, because it essentially gives MPC more choice on which sequence's first action to execute. It also reduces variance.

Plot

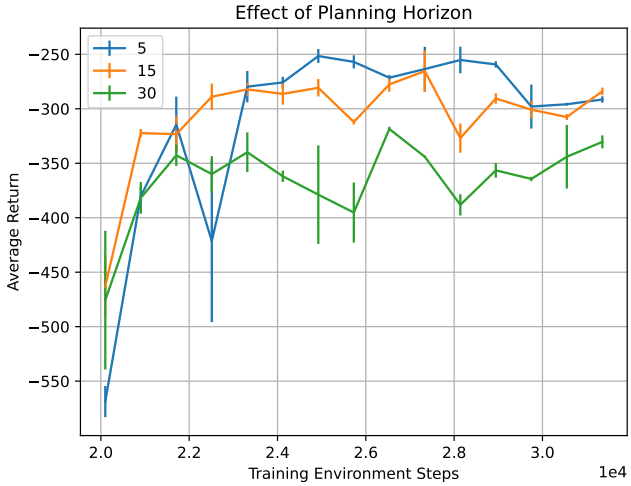


Figure 10: Effect of planning horizon is not so significant, because choosing from a small set of sequence with similar performance won't affect much, and we are only executing the first action of a sequence.

5 Problem 5: Hyper-parameter Comparison (Bonus)

Plot

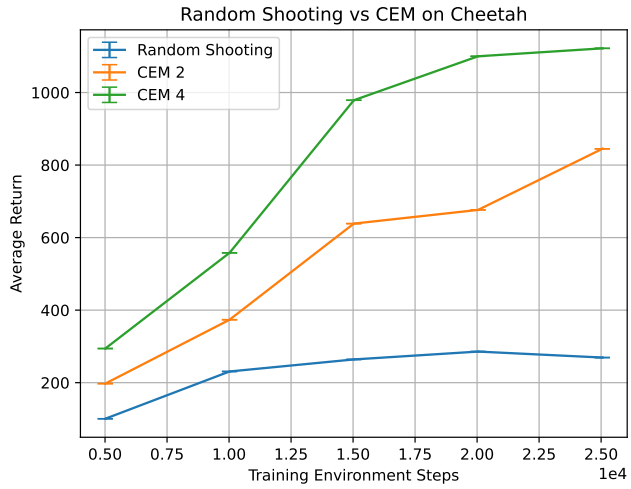
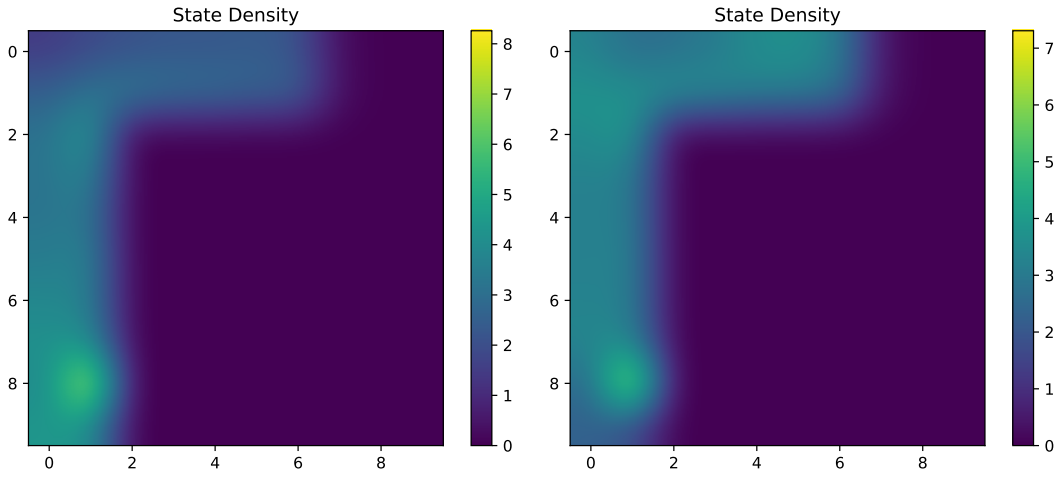


Figure 11: Average Evaluation Return comparison between Random Shooting and Cross Entropy Method.

Gaussian means and variances are modeled done to the very bottom. CEM performs much better than RS with even a few iterations. More iterations improves action sampling on a temporarily fixed MPC policy, there's still plenty of room for improvement on hyper parameters.

6 Problem 6: Exploration (Bonus)

Plot



(a) Random. (b) RND.

Figure 12: Final state density in Point Mass Easy Environment.

Plot

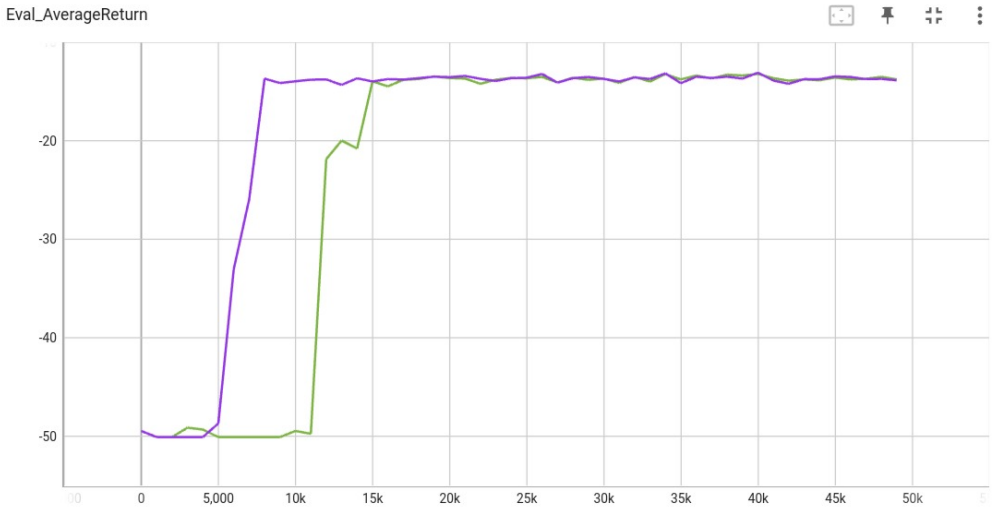
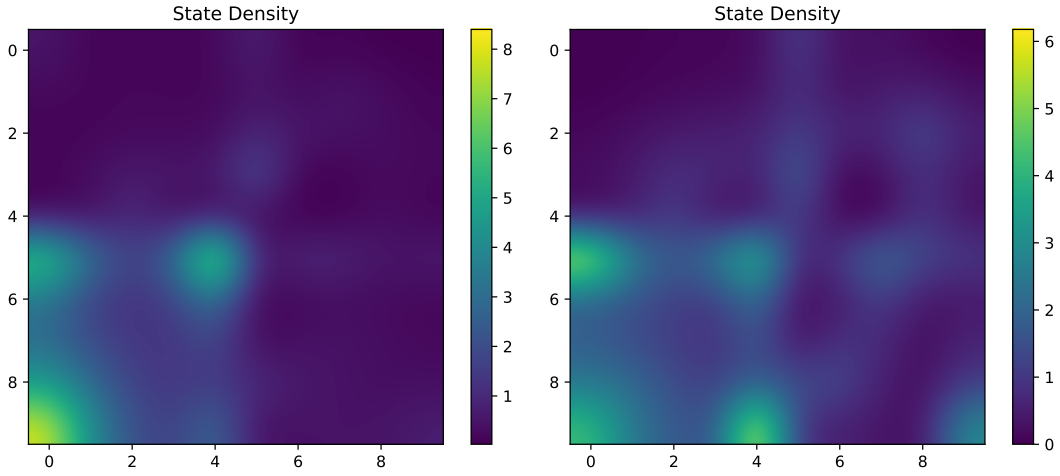


Figure 13: Average Evaluation Return in Point Mass Easy Environment. Purple: RND. Green: Random.

Plot



(a) Random. (b) RND.

Figure 14: Final state density in Point Mass Hard Environment.

Plot

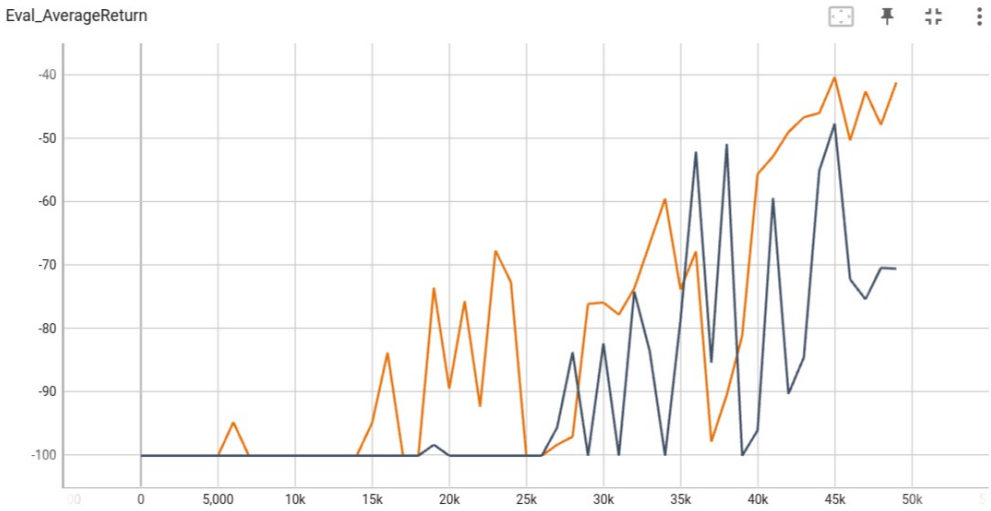


Figure 15: Average Evaluation Return in Point Mass Hard Environment. Orange: RND. Gray: Random.